

# APPROACHES TO LANGUAGE: DATA, THEORY, AND EXPLANATION

EDITED BY: Ángel J. Gallego and Aritz Irurtzun

PUBLISHED IN: Frontiers in Psychology and Frontiers in Communication





# frontiers

## Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-668-6

DOI 10.3389/978-2-88963-668-6

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# APPROACHES TO LANGUAGE: DATA, THEORY, AND EXPLANATION

Topic Editors:

**Ángel J. Gallego**, Autonomous University of Barcelona, Spain

**Aritz Irurtzun**, Centre National de la Recherche Scientifique (CNRS), France

The study of language has changed substantially in the last decades. In particular, the development of new technologies has allowed the emergence of new experimental techniques which complement more traditional approaches to data in linguistics (like informal reports of native speakers' judgments, surveys, corpus studies, or fieldwork). This move is an enriching feature of contemporary linguistics, allowing for a better understanding of a phenomenon as complex as natural language, where all sorts of factors (internal and external to the individual) interact (Chomsky 2005).

This has generated some sort of divergence not only in research approaches, but also in the phenomena studied, with an increasing specialization between subfields and accounts. At the same time, it has also led to subfield isolation and methodological a priori, with some researchers even claiming that theoretical linguistics has little to offer to cognitive science (see for instance Edelman & Christiansen 2003). We believe that this view of linguistics (and cognitive science as a whole) is misguided, and that the complementarity of different approaches to such a multidimensional phenomenon as language should be highlighted for convergence and further development of its scientific study (see also Jackendoff 1988, 2007; Phillips & Lasnik 2003; den Dikken, Bernstein, Tortora & Zanuttini 2007; Sprouse, Schütze & Almeida 2013; Phillips 2013).

**Citation:** Gallego, Á. J., Irurtzun, A., eds. (2020). Approaches to Language: Data, Theory, and Explanation. Lausanne: Frontiers Media SA.  
doi: 10.3389/978-2-88963-668-6

# Table of Contents

- 05** *Editorial: Approaches to Language: Data, Theory, and Explanation*  
Ángel J. Gallego and Aritz Irurtzun
- 07** *Influence of Perceptual Saliency Hierarchy on Learning of Language Structures: An Artificial Language Learning Experiment*  
Tao Gong, Yau W. Lam and Lan Shuai
- 17** *Reversing the Approach to Null Subjects: A Perspective From Language Acquisition*  
Maia Duguine
- 28** *Dissociating Effects of Scrambling and Topicalization Within the Left Frontal and Temporal Language Areas: An fMRI Study in Kaqchikel Maya*  
Shinri Ohta, Masatoshi Koizumi and Kuniyoshi L. Sakai
- 42** *Grammatical Role Parallelism Influences Ambiguous Pronoun Resolution in German*  
Antje Saueremann and Natalia Gagarina
- 50** *Understanding Grammars through Diachronic Change*  
Nerea Madariaga
- 66** *Syntactic Priming as a Test of Argument Structure: A Self-paced Reading Experiment*  
Isabel Oltra-Massuet, Victoria Sharpe, Kyriaki Neophytou and Alec Marantz
- 80** *On the Nature of Clitics and Their Sensitivity to Number Attraction Effects*  
Mikel Santesteban, Adam Zawiszewski, Kepa Erdocia and Itziar Laka
- 101** *Toward Cognitively Constrained Models of Language Processing: A Review*  
Margreet Vogelzang, Anne C. Mills, David Reitter, Jacolien Van Rij, Petra Hendriks and Hedderik Van Rij
- 119** *Strong Generative Capacity and the Empirical Base of Linguistic Theory*  
Dennis Ott
- 124** *The Locus Preservation Hypothesis: Shared Linguistic Profiles Across Developmental Disorders and the Resilient Part of the Human Language Faculty*  
Evelina Leivada, Maria Kambanaros and Kleanthes K. Grohmann
- 138** *The Relationship Between Syntactic Satiation and Syntactic Priming: A First Look*  
Monica L. Do and Elsi Kaiser
- 157** *On the Diversity of Linguistic Data and the Integration of the Language Sciences*  
Roberta D'Alessandro and Marc van Oostendorp
- 161** *Sentence Repetition as a Tool for Screening Morphosyntactic Abilities of Bilingual Children With SLI*  
Elena Theodorou, Maria Kambanaros and Kleanthes K. Grohmann
- 174** *Length of Utterance, in Morphemes or in Words?: MLU3-w, a Reliable Measure of Language Development in Early Basque*  
Maria-José Ezeizabarrena and Iñaki Garcia Fernandez

- 191** *Backward Dependencies and in-Situ wh-Questions as Test Cases on How to Approach Experimental Linguistics Research That Pursues Theoretical Linguistics Questions*  
Leticia Pablos, Jenny Doetjes and Lisa L.-S. Cheng
- 208** *The Limited Role of Number of Nested Syntactic Dependencies in Accounting for Processing Cost: Evidence From German Simplex and Complex Verbal Clusters*  
Markus Bader
- 229** *Scalar and Ignorance Inferences are Both Computed Immediately Upon Encountering the Sentential Connective: The Online Processing of Sentences With Disjunction Using the Visual World Paradigm*  
Likun Zhan
- 240** *ULTRA: Universal Grammar as a Universal Parser*  
David P. Medeiros
- 254** *Assessing the Role of Experimental Evidence for Interface Judgment: Licensing of Negative Polarity Items, Scalar Readings, and Focus*  
Anastasia Giannakidou and Urtzi Etxeberria
- 272** *Handling Sign Language Data: The Impact of Modality*  
Josep Quer and Markus Steinbach



# Editorial: Approaches to Language: Data, Theory, and Explanation

Ángel J. Gallego<sup>1\*</sup> and Aritz Irurtzun<sup>2\*</sup>

<sup>1</sup> Department of Spanish Philology, Autonomous University of Barcelona, Barcelona, Spain, <sup>2</sup> CNRS, IKER (UMR 5478), Bayonne, France

**Keywords:** language, data, theory, description, analysis

## Editorial on the Research Topic

### Approaches to Language: Data, Theory, and Explanation

This Research Topic serves as a showroom for the latest developments in linguistic methods and approaches. In so doing, the articles go beyond developing a specific research problem and they also serve as a sample of the kind of methods employed in different approaches to language, in the hope that this discussion prompts a reflection on the relation between theory, data, evidence, and explanation.

Madariaga's article is a clear vindication of the role of different factors shaping languages. It takes an I-language perspective in order to explain certain phenomena that are otherwise unapproachable such as the variation in object case marking of several Russian verbs.

Ezeizabarrena and Garcia Fernandez analyze the feasibility and utility of words or morphemes as measures for (morpho-)syntactic development in agglutinative languages such as Basque, confirming their reliability for identifying developmental patterns.

Theodorou et al. provide a pioneering analysis of sentence repetition tasks as useful tools for assessing children's language ability in bilingual settings. The study validates the diagnostic accuracy of the task, showing that it has the potential to be used as a referral criterion to identify children with Specific Language Impairment (SLI).

Leivada et al. advance in the demarcation of the linguistic phenotype of three developmental disorders: SLI, Down syndrome, and autism spectrum disorder. They perform a systematic and cross-linguistic review of their linguistic profiles and formulate the Locus Preservation Hypothesis, suggesting that aspects of the language faculty are immune to impairment across developmental disorders.

Various experimental studies address the processing of long-distance dependencies. Santesteban et al. study whether antecedent-clitic dependencies in Spanish are computed like agreement or like pronominal dependencies. They report two experiments arguing for cue-retrieval accounts of dependency resolution and suggesting that the sensitivity to attraction effects shown by clitics resembles more the computation of pronominal dependencies than that of agreement. Likewise, Sauermann and Gagarina report a visual world eye-tracking study investigating the impact of the word order and grammatical role parallelism on the online comprehension of pronouns in German. It provides compelling evidence that pronouns may not in general be associated with the subject or topic of a sentence but that their resolution is modulated by additional factors.

In a different setting, Pablos et al. present some experiments on the processing of long-distance backward dependencies in Dutch and the processing of *in-situ* wh-questions in Mandarin vs. French. This is also a study that provides a general reflection on the challenges that experimental work faces in finding a compromise between addressing theoretically relevant questions and being able to implement them in a controlled experimental paradigm.

## OPEN ACCESS

### Edited and reviewed by:

Yury Y. Shtyrov,  
Aarhus University, Denmark

### \*Correspondence:

Ángel J. Gallego  
angel.gallego@uab.cat  
Aritz Irurtzun  
aritz.irurtzun@iker.cnrs.fr

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 25 June 2020

**Accepted:** 15 September 2020

**Published:** 22 October 2020

### Citation:

Gallego AJ and Irurtzun A (2020)  
Editorial: Approaches to Language:  
Data, Theory, and Explanation.  
Front. Psychol. 11:576244.  
doi: 10.3389/fpsyg.2020.576244

On the more theoretical prism, Medeiros centers on design properties of language, proposing a Universal Linear Transduction Reactive Automaton (ULTRA) directly mapping surface word orders to underlying base structure with a stack-sorting algorithm.

Taking a historical and epistemological stance, Ott asks for a paced consideration of the implications of “strong generativity” in the field, and its relation to data, judgments, and their relationship with theoretical evidence.

D’Alessandro and van Oostendorp provide an even bigger picture by addressing directly broad ontological questions about our object of study and epistemological questions about how to best study it. The position that they defend is a plural one, vindicating the necessity of different disciplines, views and methodologies when studying language.

Quer and Steinbach analyze the impact of modality on linguistic data elicitation and collection, corpus studies, and experimental studies highlighting a set of specific challenges for sign language research. This paper also vindicates the complementarity of theoretical approaches and experimental studies.

Duguine proposes a new model for null subjects, and focuses on its implications for language development. The paper explores the consequences of an inverse approach to pro-drop in the domain of language acquisition, arguing that it allows to account for a number of properties of child languages.

Do and Kaiser analyze syntactic satiation effects. Their experimental analysis of Subject island and Complex-NP Constraint violations uncovers different factors that may bring about satiation, and the overall conclusion is that satiation may not be a one-size-fit-all phenomenon for different types of structures.

Bader centers on the processing of center embedding constructions in German. As a result of the discussion of the three novel experiments he reports, he argues for a multifactorial account of the limitations on center embedding in natural languages.

Giannakidou and Etxeberria review a series of experimental studies that address complex judgments involving integration from multiple levels of grammatical representation. They show the welcome results of the combination of theoretical research and experimental techniques when addressing such complex phenomena as NPI licensing or the emergence of scalar readings.

Zhan also provides a nice example of the usefulness of experimental methods (eye-tracking) when addressing questions such as how and when scalar and ignorance inferences are computed in disjunction phrases.

Ultra-Massuet et al. present the results of a structural priming experiment where they test two different theoretical approaches to the argument structure of (in)transitive structures. The study suggests a stronger predictive contribution of a model that supports an interpretive semantics view of syntax.

Vogelzang et al. analyze how language processing interacts with general cognitive resources by reviewing different language processing models.

Ohta et al. explore the hypothesis that topicalization and scrambling constructions are quite different in nature. They set up an experiment to assess the modular nature of these structures in Kaqchikel Mayan by identifying their main processing loci.

Finally, Gong et al. report an artificial language learning experiment studying whether hierarchies in perceptual saliency influence the learning of orders regulating adjectives of involved visual features. Their results show learning biases for orders that are congruent with the perceptual saliency hierarchy, which could contribute to the structural configuration of languages.

In a nutshell, this Research Topic offers a wide panoramic view of different stances and approaches to language and shows how the interaction of a robust theoretical apparatus, plus the application of cutting-edge data acquisition and analysis techniques can help us move forward in the understanding of a phenomenon as complex and poliedric as natural language.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This research has been partially supported by grants from the Ministerio de Economía y Competitividad (FFI2017-87140-C4-1-P; PGC2018-096870-B-I00), the Generalitat de Catalunya (2017SGR634), the Institució Catalana de Recerca i Estudis Avançats (ICREA Acadèmia 2015), and the Agence Nationale de la Recherche (UV2 ANR18-FRAL0006 (ANR-DFG); BIM ANR-17-CE27-0011).

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Gallego and Irurtzun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Influence of Perceptual Saliency Hierarchy on Learning of Language Structures: An Artificial Language Learning Experiment

Tao Gong<sup>1,2\*</sup>, Yau W. Lam<sup>3</sup> and Lan Shuai<sup>1</sup>

<sup>1</sup> Haskins Laboratories, New Haven, CT, USA, <sup>2</sup> Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, China, <sup>3</sup> Department of Linguistics, University of Hong Kong, Hong Kong, China

Psychological experiments have revealed that in normal visual perception of humans, color cues are more salient than shape cues, which are more salient than textural patterns. We carried out an artificial language learning experiment to study whether such perceptual saliency hierarchy (color > shape > texture) influences the learning of orders regulating adjectives of involved visual features in a manner either congruent (expressing a salient feature in a salient part of the form) or incongruent (expressing a salient feature in a less salient part of the form) with that hierarchy. Results showed that within a few rounds of learning participants could learn the compositional segments encoding the visual features and the order between them, generalize the learned knowledge to unseen instances with the same or different orders, and show learning biases for orders that are congruent with the perceptual saliency hierarchy. Although the learning performances for both the biased and unbiased orders became similar given more learning trials, our study confirms that this type of individual perceptual constraint could contribute to the structural configuration of language, and points out that such constraint, as well as other factors, could collectively affect the structural diversity in languages.

**Keywords:** perceptual saliency hierarchy, artificial language learning, syntax, learning bias, diversity

## OPEN ACCESS

### Edited by:

Ángel J. Gallego,  
Autonomous University of Barcelona,  
Spain

### Reviewed by:

Cristiano Chesì,  
IUSS Pavia, Italy  
María Del Carmen Horno-Chéliz,  
University of Zaragoza, Spain

### \*Correspondence:

Tao Gong  
gong@haskins.yale.edu

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 28 September 2016

**Accepted:** 29 November 2016

**Published:** 21 December 2016

### Citation:

Gong T, Lam YW and Shuai L (2016)  
Influence of Perceptual Saliency  
Hierarchy on Learning of Language  
Structures: An Artificial Language  
Learning Experiment.  
Front. Psychol. 7:1952.  
doi: 10.3389/fpsyg.2016.01952

## INTRODUCTION

Physical objects can be discriminated by visual features such as color, shape, and texture. Human eyes are essentially light receptors, and thus, color or brightness information requires little cognitive load for processing, thus becoming the strongest cue for visual perception. In terms of evolution, the alimentary “niche” also enhanced color perception in humans and other primates (Dominy and Lucas, 2001; Melin et al., 2012). Difference in color or brightness enables humans to perceive additional features such as shape and textural pattern. Per these fundamental features (color, shape, and texture), psychological experiments have explicitly shown that: random variations in color interfere with viewer’s ability to identify shapes, but variations in shape have no explicit effects (in terms of judgement accuracies and reaction times) on color discrimination (Callaghan, 1990; Healey, 2000); and random variations in color or shape interfere with viewer’s identification of visual patterns of texture, but not vice-versa (Treisman, 1985; Healey and Enns, 1999). This evidence reveals a *perceptual saliency hierarchy* (PSH, the relative conspicuousness of various visual features at first exposure, Healey and Enns, 2012), which states that in normal visual



perception of humans, color information appears to be more salient than shape information, and shape more salient than visual textural pattern (simply, color > shape > texture).

Language serves as the primary means for humans to describe visual features. Given the PSH, an interesting question arises: Whether the PSH can cast any influence on learning or processing the language structures used to regulate the relevant adjectives of those visual features. Answer to this question helps reveal the relationship between structural configuration in language and perceptual or cognitive constraints in humans, which is a challenging issue in modern psychology and linguistics (Christiansen and Kirby, 2003; Gentner and Goldin-Meadow, 2003; Hurford, 2007, 2012).

Many approaches have been adopted to study this issue. Corpus analyses have identified universal characteristics in language structures and potential links between language structures and cognitive constraints in humans (Ferrer-i-Cancho, 2004; Liu, 2010; Futrell et al., 2015). Computational modeling (Gong et al., 2013, 2014) has demonstrated how psychological or physiological constraints help shape word order (Gong et al., 2009), compositionality (Kirby, 1999; Brighton, 2002; Smith et al., 2003; Kirby et al., 2006), and syntactic patterns such as recursion, case, or long-distance dependency (Elman, 1990; Conway and Christiansen, 2001; Christiansen and Ellefson, 2002; Lupyán and Christiansen, 2002; Reali and Christiansen, 2009; Christiansen and Chater, 2015). In particular, some simulations show that the word order bias (in favor of certain orders like SOV or SVO but against others like VOS or OVS) in the world's languages could result from individual perceptual constraint, which takes effect during communications (Gong et al., 2009). Other simulations illustrate that the universal color naming patterns in the world's languages could result from the perceptual constraint of human eyes towards colors, which also takes effect during cultural transmission of color terms (Baronchelli et al., 2012). These studies have illustrated the effect of perceptual or cognitive constraints on structural configuration of language (Heine and Kuteva, 2008; Chater et al., 2009; Mesoudi, 2011; Richerson and Christiansen, 2013).

In experimental psychology, the paradigm of artificial language learning (ALL, in which participants are asked to learn a language or language-like system, and then tested on what they have learned; depending on underlying structures, ALL is also called artificial grammar learning) has been used to investigate issues concerning language and cognition (Esper, 1925; Reber, 1967; Folia et al., 2010; Onnis, 2012). An ALL experiment typically consists of a sequence of learning (a.k.a. training) and testing phases, which alternate throughout the experiment. In a learning phase, participants are presented with visual or auditory symbols concatenated following a predefined grammar-like structure. In the subsequent testing phase, they are presented with already-seen or unseen instances. Individual learning is said to occur when participants can distinguish instances that respect the underlying structure from those that violate it.

Artificial language learning experiments can design pseudo-words and structures distinct from participants' native language to diminish the influence of participants' prior linguistic knowledge and highlight corresponding learning mechanisms

and factors hard to control in naturalistic scenarios (Onnis, 2012). They can also generate sufficient instances to trace individual learning and evaluate whether individuals can generalize their learned knowledge to unseen instances. In addition, by recruiting human participants and carefully designed artificial languages, ALL experiments can complement other approaches, such as verifying simulated behaviors and modeling results to bridge the gap between language processing in humans and relevant mechanisms in artificial agents (Kirby et al., 2008; Cornish, 2010). Furthermore, it has been repetitively shown that ALL experiments can uncover the same (or similar) mechanisms manifest in natural and artificial language processing (Reber, 1993; Gómez and Gerken, 2000; Pothos, 2007) and in first (Misyak et al., 2010; Misyak and Christiansen, 2012) and second language acquisition (Friederici et al., 2002; Robinson, 2010; Brooks et al., 2011; Petersson et al., 2012; Morgan-Short et al., 2014; Ettlínger et al., 2015). These advantages have made ALL experiments revitalize language learning research in the past century (Braine, 1963; Moeser and Bregman, 1972; Saffran et al., 1996; Morgan and Newport, 1981; Tily et al., 2011; Tabullo et al., 2012).

To our knowledge, there are no modeling or experimental studies that address directly the PSH and its influence on language learning. In this paper, we conducted an ALL experiment to study this issue. A number of artificial languages were designed, each describing two out of the three types of visual features in the PSH. In an artificial language, a visual feature was mapped to a phonetic segment, and segments, respectively, encoding the two features followed a consistent order. We referred to the theme-first principle in linguistics to clarify such orders. The principle states that more "thematic" information tends to precede less "thematic" one in normal linguistic expressions (Tomlin, 1986) (here, thematic information refers to the pragmatic or psycholinguistic reflex of the general attention in human cognition). This principle helps account for many cross-language phenomena, especially for word order (Halliday, 1967; Mathesius, 1975; Lambrecht, 1994; Cinque, 1999, 2010; Cinque and Rizzi, 2008; Longobardi and Guardiano, 2009). In terms of visual perception, it suggests that information of perceptually more salient feature should precede that of less salient feature. Following this principle, in our ALL experiment, we regarded an order as *congruent* with the PSH, if it puts a segment encoding a perceptually more salient feature in front of a segment encoding a less salient feature; otherwise, the order is deemed *incongruent*. We recruited human participants to learn, by repetitive exposure of instances, the artificial languages having congruent or incongruent orders, and assessed individual learning using seen and unseen instances.

In the following sections, we described the experiment, reported its results, and discussed the relation between language and human cognition based on this study.

## MATERIALS AND METHODS

The experimental protocol was approved by the College Research Ethics Committee of University of Hong Kong. The methods

were carried out in accordance with the approved guidelines from the College Research Ethics Committee. Informed consents were obtained from all participants.

### Participants

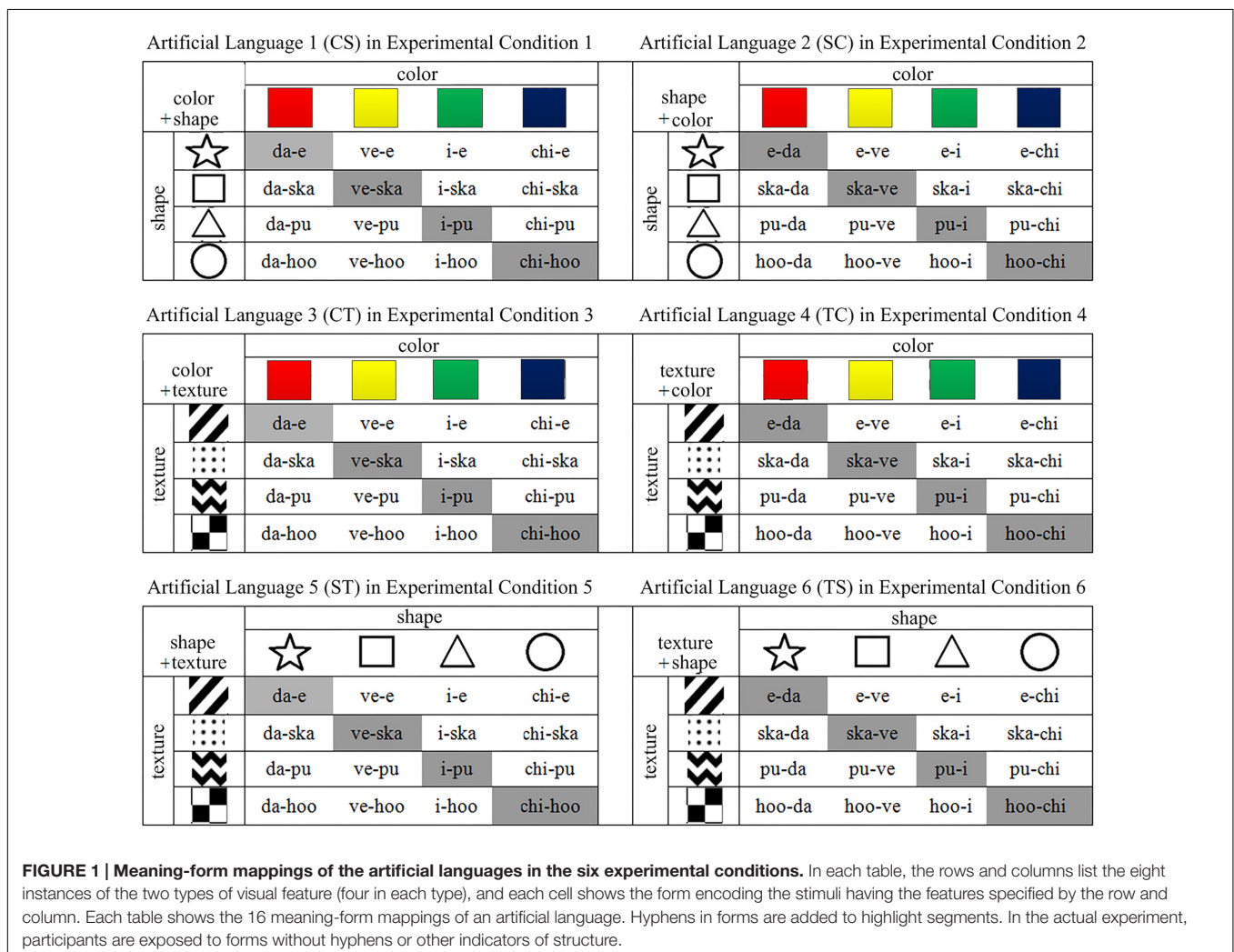
One hundred and thirty-two students from the University of Hong Kong participated into the experiment (66 females, mean age = 23.27, age range = 17–33, *SD* = 4.97). Forty Hong Kong dollars were paid to participants who had finished the experiment and filled in the post-experiment surveys. The recruited participants had normal or adjusted-to-normal vision, and reported no history of developmental delay or acquired neurological disorder. They were native Mandarin or Cantonese speakers, and had an intermediate level of English.

### Materials

We defined six artificial languages, respectively, used in six experimental conditions (see **Figure 1**). Each artificial language described two of the three visual features in the PSH (color (C) and shape (S), color and textural pattern (T), and shape and textural pattern). There are two reasons for considering orders

between only two visual features. First, as shown in previous psychological experiments of visual feature saliency (Treisman, 1985; Callaghan, 1990; Healey and Enns, 1999; Healey, 2000), using languages encoding only two visual features can directly reflect whether the congruent or incongruent orders between the two affect the learning of those orders. Second, training participants on orders among three visual features would require more learning trials to give participants enough opportunities to detect and learn similarities between visual features and segments and similarities in regulating orders among segments. This would increase learning difficulty and memory burden, extend experiment time, and might have adverse effects on participants' motivation.

Our training stimuli consisted of 48 images created by PhotoImpact X3. Each image depicted an object with a unique combination of a shape (star, square, triangle, or circle), a color (red, yellow, green, or blue), and a textural pattern (stripes, dots, zigzag, or checkerboard). These images were divided evenly into three sets. Each set of 16 images differed in two features (e.g., color and shape) and were the same in the third (e.g., texture) (see **Figure 1**).



All forms of the artificial languages were presented visually in the experiment. A form consisted of two compositional segments. A segment encoded one instance of a visual feature and had a consonant-vowel or vowel structure. All segments had roughly the same level of learning difficulty, and did not resemble any orthography of real words in English or any pronunciation of real characters in Mandarin or Cantonese. We also designed the segments to avoid iconicity (perceptuomotor analogies between aspects of a form and meaning of a word, e.g., onomatopoeia words and ideophones, Dingemans, 2012; Dingemans et al., 2015), which could assist language learning or comprehension (Simner et al., 2010; Pemiss and Vigliocco, 2014). In each form of an artificial language, the two segments followed a consistent order. Depending on encoded visual features, the order between segments was either congruent or incongruent with the PSH.

As shown in **Figure 1**, languages 1 and 2 describe color (C) and shape (S), languages 3 and 4 described color (C) and texture (T), and languages 5 and 6 described S and T. Each pair of the languages were formed by the same set of segments but differed in regulating order. Three of these languages (languages 1, 3, and 5) had congruent orders (CS, CT, and ST), and the other three (languages 2, 4, and 6) had incongruent orders (SC, TC, and TS).

## Procedure

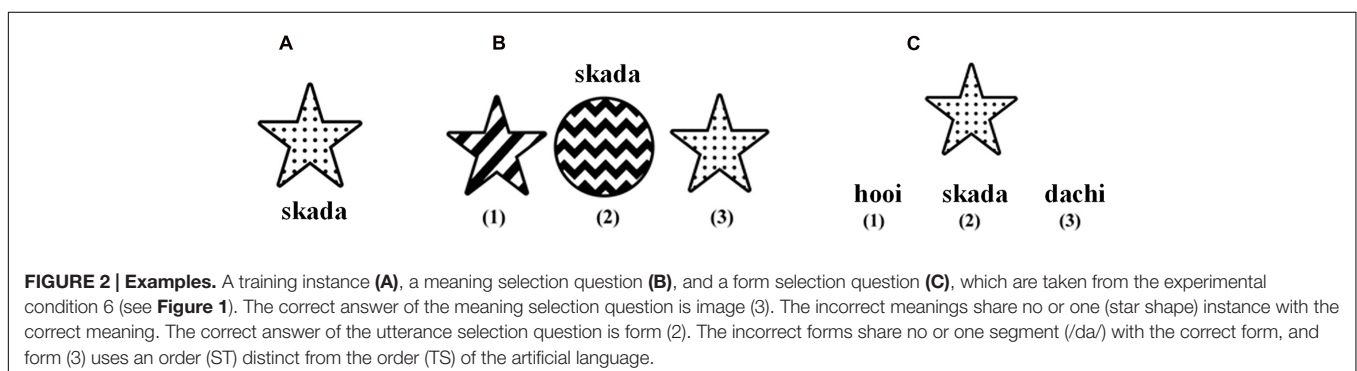
The procedure was implemented using E-Prime 2.0. During the experiment, participants sat comfortably in front of a laptop in a bright, quiet room. They were asked to learn an “alien language” by viewing its meaning-form mappings displayed on a 21-inch computer monitor at a resolution of 1280 × 1024 and a refresh rate of 75 Hz. The font size was 64 pixels. The distance between the screen and participants’ eyes was approximately 64 cm. We used a between-subject design; each participant was assigned to one experimental condition to learn the corresponding artificial language. Gender and number of participants were balanced in each condition (11 females and 11 males). Prior to the experiment, the participants went through a two-minute familiarization block.

The experiment consisted of three 5-min blocks, with optional two-minute breaks in between; the whole experiment lasted about 20 min. A block consisted of a learning and a testing phase; in total, there were three learning and three testing phases to trace learning progress.

In a learning phase, 12 out of the 16 meaning (image)-form mappings (those in the white cells in **Figure 1**) of an artificial language were displayed visually to participants. A mapping was shown on the center of the monitor, with the form presented simultaneously underneath the image (see **Figure 2A**). A mapping remained visible for five seconds. Presentation of all 12 mappings was repeated three times. Each time the meaning-form pairs were displayed in a pseudo-random order ensuring that the images of any two consecutively presented mappings shared no instances of the two types of visual features that the artificial language described. This setting prevented the participants from immediately noticing the associations between the visual features and the segments, thus increasing the difficulty of the learning task.

In a testing phase, individual learning was assessed by 20 forced-choice questions presented in a pseudo-random order. Participants gave their answers by key pressing. After the participants answered a question, the next one popped up without feedback. Ten of the questions were meaning selection questions (see **Figure 2B** for an example). In each of them, the participants saw a form followed by three meanings (images) displayed in a pseudo-random order. Participants were asked to select the image that they believed was expressed by the form. Incorrect meanings shared at most one instance of the visual feature with the correct one. The other ten questions were form selection questions (see **Figure 2C** for an example). In each of them, one image and three forms were displayed simultaneously. The participants were asked to select the form that they believed encoded the meaning. Incorrect forms shared at most one segment with the correct form. The segment orders in the incorrect forms were distinct from the order used in the instances in the learning phase.

The 12 meaning-form mappings shown in the learning phases appeared at least once and at most twice as the correct answers in the 20 testing questions. Each mapping had the same occurrence frequency in the learning and testing phases. To answer the testing questions correctly, the participants needed to learn not only the mappings between the visual features and the segments but also the order between the segments. Compared with the much larger search space in the free recall tasks as in previous studies (e.g., Cornish, 2010; Tamariz and Kirby, 2015), answers in the forced choice questions were more



**FIGURE 2 | Examples.** A training instance (A), a meaning selection question (B), and a form selection question (C), which are taken from the experimental condition 6 (see **Figure 1**). The correct answer of the meaning selection question is image (3). The incorrect meanings share no or one (star shape) instance with the correct meaning. The correct answer of the utterance selection question is form (2). The incorrect forms share no or one segment (/da/) with the correct form, and form (3) uses an order (ST) distinct from the order (TS) of the artificial language.

limited and allowed explicitly tracing the participants' learning performances.

In the last testing phase, apart from the 20 normal testing questions containing the items already seen in the learning phases, there were additional four meaning selection and four form selection questions that contained the novel meaning-form mappings not presented in the learning phases (those in the gray cells in **Figure 1**). Performance on these items helped evaluate whether the participants could generalize their learned knowledge to unseen instances. All the 28 questions were presented in a pseudo-random order.

## Measures

In each testing phase, we recorded each participant's accuracy (percentage of correct answers to the questions of the same type) and average reaction time to each of the meaning and form selection questions. In the last testing phase, apart from the measures to the normal testing questions, we also recorded the accuracies and average reaction times to the additional questions about the novel items. We grouped the accuracy and average reaction time data in the experimental conditions 1, 3, and 5 (the artificial languages therein had congruent orders) as the congruent set, and those in the conditions 2, 4, and 6 as the incongruent set. In each set, the accuracies and average reaction times were grouped according to the three testing phases. The measures to the additional questions formed the fourth phase. To meet the assumption of normality, we used the log-transformed (base  $e$ ) reaction times in the analyses.

After the experiment, the participants were asked to fill a *post hoc* survey to indicate: which type of questions – meaning or form selection – was harder to answer; in which block they could confidently learn the “alien language”; and how difficult they felt to learn the “alien language” on a scale of 1 to 5, ‘1’ being the easiest, ‘3’ being neutral, and ‘5’ being the hardest.

## Preprocessing and Analyses

Following the general procedure in assessing experimental data (Osborne and Overbay, 2004), we removed the outliers from the accuracy and reaction time data before the analyses. Outliers were values exceeding 2.5 standard deviations from the group mean. For accuracies, outliers were accuracies that were too low; for reaction times, they were times either too long or too short. Among the 1056 ( $132 \times 4 \times 2$ ) accuracy data in eight groups, 34 outliers were removed; for the reaction times, 23 were removed. Another way to handle outliers is to replace them with the group

means, the results following this procedure were similar (see Supplementary Table S1).

We conducted two ANOVAs, respectively, on accuracy and average reaction time to test our working hypothesis that the PSH affects the learning of regulating orders between segments encoding the involved visual features. In the ANOVAs, we treated the congruency of artificial languages as a between-subjects factor (two levels: the congruent languages, those used in conditions 1, 3, and 5, and incongruent languages, those used in conditions 2, 4, and 6), and the experimental phase as a within-subjects factor (four levels: the testing phases 1, 2, 3, and 4, the latter of which consists of the measures to the additional testing questions involving the unseen items). The ANOVA tests also took into account the question type (two levels: meaning selection or utterance selection) and interaction between congruency and experimental phase. In addition to the ANOVA tests, we conducted group t-tests to compare the accuracies and average reaction times between the conditions differing in regulating orders (conditions 1 vs. 2, 3 vs. 4, and 5 vs. 6), which aimed to reveal possible learning biases for the congruent or incongruent orders. Following the Bonferroni correction, we set the critical  $p$  value to identify significant effects as 0.002 ( $0.05/(2+24)$ ), 26 tests in total). All the analyses were carried out in R 3.2.4 (R Core Team, 2016).

## RESULTS

**Table 1** shows the results of the ANOVAs. In both tests, question type showed no significant effect, which matched the *post hoc* surveys; 125 participants felt invariant to both types of questions. This indicated that the way of recording the individual learning performance in our study had no obvious effect on the recorded results.

Both congruency and experimental phase showed significant main effects, but there was no significant interaction between the two. Compared with experimental phase, congruency had a smaller effect size  $\eta^2$ . The significant effect of congruency confirmed our working hypothesis that the perceptual saliency hierarchy could affect individual learning of congruent or incongruent orders. The significant effect of experimental phase indicated that learning occurred at different experimental phases of instance exposure. The non-significant interaction between congruency and experimental phase suggested that the learning patterns across phases for the congruent and incongruent orders were largely the same.

**TABLE 1 | Results of the ANOVAs of accuracy (ACC) and average reaction time (RT).**

Factor	ANOVA of ACC			ANOVA of RT		
	<i>F</i>	<i>P</i>	$\eta^2$	<i>F</i>	<i>p</i>	$\eta^2$
Congruency	<b>30.413</b>	<b>&lt;0.00005</b>	<b>0.029</b>	<b>17.442</b>	<b>&lt;0.00005</b>	<b>0.017</b>
Phase	<b>40.006</b>	<b>&lt;0.00001</b>	<b>0.106</b>	<b>24.090</b>	<b>&lt;0.00001</b>	<b>0.066</b>
Question Type	0.819	0.366		0.558	0.455	
Congruency × Phase	4.439	0.004		2.867	0.036	

Significant effects (whose  $p$ -values are below the critical  $p$ -value 0.002) are highlighted in bold.

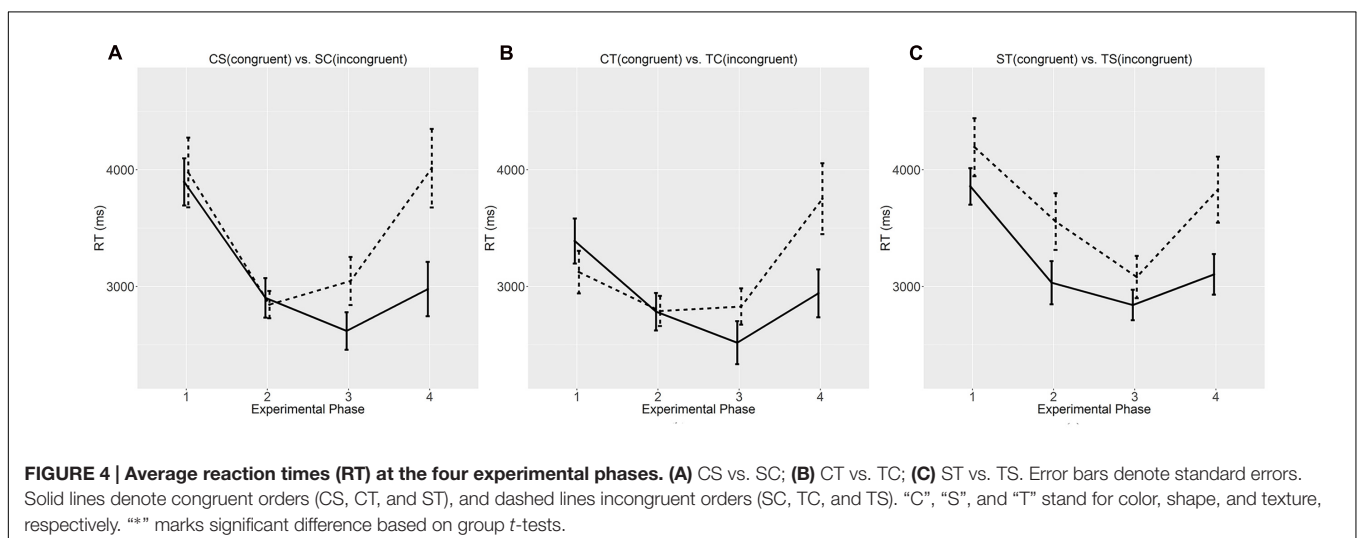
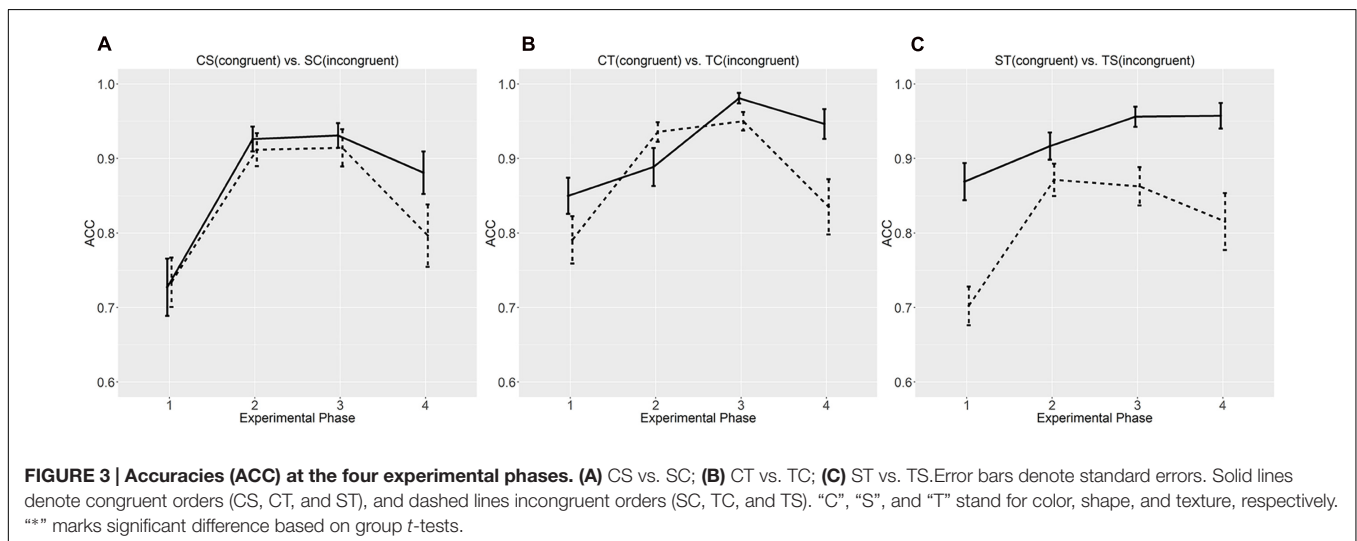
**Figures 3 and 4** compare the accuracies and average reaction times across the four phases between the conditions differing in regulating orders. Across the three phases, there were a general increase in accuracy and a general decrease in average reaction time, which echoed the significant effect of experimental phase in the ANOVAs. In our experiment, individual learning started at the first phase, and the improvement after the third phase was smaller than that after the second phase. After three phases, the participants had largely grasped the compositional languages in different conditions. These results also matched the *post hoc* surveys; 128 participants claimed that they had confidently learned most of the meaning-form mappings after the second phase, and the other four said that they had learned the language right after the first phase.

Although the participants claimed to have learned the artificial languages, their performances on the unseen items at the last phase revealed some biases for the congruent orders. As for the CS and SC orders, the participants showed similar accuracies, but their reaction times to the CS orders were shorter than those to

the SC orders. As for the CT and TC orders, they showed higher accuracies and shorter reaction times to the CT order than the TC order. As for the ST and TS orders, they showed a similar bias for the ST order over the TS order. Significant difference in accuracy was also shown at the first and third testing phases concerning the seen instances. These biases also manifested in the *post hoc* surveys; when asked to evaluate the learning difficulty of the “alien language”, the average scores given by the participants in the CS and SC conditions were similar (2.38 vs. 2.64), but those in the CT and ST conditions were different (2.86 vs. 3.65), so were their scores in the ST and TS conditions (3.10 vs. 3.95).

## DISCUSSION

In this paper, we evaluated whether the perceptual constraint regarding the saliency hierarchy of the basic visual features affects the learnability of ordering structures between the segments encoding such features in an artificial language. After repeated



exposure to the tokens of the artificial languages with different orderings, the participants gradually learned the segments encoding color, shape, or textural patterns and the orders between these segments. Their judgements on the unseen instances indicated that they could generalize their learned knowledge and apply it to novel items. Moreover, they exhibited biases for the orders that were congruent with the perceptual saliency hierarchy regarding color, shape, and textural patterns. To be specific, they showed strong biases for the CT (color before textural pattern) and ST (shape before textural pattern) orders over the TC and TS orders, in terms of judging accuracy and average reaction time. Such biases started to exhibit during the learning process. They also showed a weak bias for the CS over the SC order, which only manifested in average reaction time when judging the unseen items.

In this ALL experiment, the observed biases were not induced by participants' prior linguistic knowledge (of Mandarin, Cantonese, or English). In simple phrases of Mandarin or Cantonese, information of textural pattern often appears before that of shape, and color before shape (e.g., "*hongse* (red) *mutou* (wood) *yuan* (round) *zhuozi* (table)") (Yip and Rimmington, 2004; Zhu, 2005), whereas the participants in our experiment exhibited a strong bias for the orders putting textural pattern after color or shape. In simple phrases of English, adjectives of shapes often appear in front of those of colors (e.g., "a round red wood table") (Carter et al., 2011), but the participants showed no bias for color and shape at least in accuracy. In addition, the participants had no previous experience of the segments used in our experiment, and had no chance to apply their prior linguistic knowledge to change the artificial languages or develop one from scratch. These ensure that the observed patterns can be safely ascribed to the perceptual saliency hierarchy. Nonetheless, we acknowledge that participants' alphabetic knowledge may potentially affect their performance in ALL. This is an inevitable limitation of ALL experiments recruiting alphabetic language speakers to learn alphabetic languages. Recruiting participants with no alphabetic experiences (e.g., pre-language children) or using uncommon symbols or non-linguistic forms to design artificial languages may help diminish such influence, as in experimental semiotics studies (Galantucci and Garrod, 2010, 2011) (e.g., Galantucci, 2005; Scott-Phillips et al., 2009; Taylor et al., 2011; De Boer and Verhoef, 2012; Claidière et al., 2014; Tamariz and Kirby, 2015). However, many of such studies focus on the emergence of a language-like communication system out of random signals, and participants therein are allowed to introduce signals that they prefer during the recall tasks.

Our fact that the perceptual saliency hierarchy affects the learning and processing of relevant language structures reveals a close relation between perceptual constraints in humans and structural configuration in language. In a linguistic form, if the ordering of segments encoding the visual features follows naturally the perceptual saliency of those features, production and comprehension of the form would be more straightforward. Then, compared with the forms having an incongruent order between those features, the accuracies of answering questions about the forms having congruent orders tend to be higher, and

the reaction times shorter. This is evident by the strong biases for the CT and ST orders over TC and TS orders.

In addition, compared with color and shape, textural pattern is much less salient, and it is shown as contrast of color or shape. Detection of such pattern occurs after detection of color or shape, and relies on detection of color or shape (Treisman, 1985; Healey and Enns, 1999). This also explains the strong bias for the CT and ST orders. By contrast, color appears to be slightly more salient than shape, which resulted in the weak bias for the CS over SC order.

All these results are in line with the perspective that perceptual constraints affect the learning (and use) of related language structures (Heine, 2001; Pullum and Scholz, 2002; Mesoudi and Whiten, 2008; Christiansen and Chater, 2015). They also suggest that difference in saliency levels of the visual features could affect the degree of bias for the congruent orders regulating those features.

Given the bias for the congruent orders, a follow-up question arises: If the perceptual saliency hierarchy affects the regulating orders of the segments encoding the involved visual features, is this structure in all languages the same, favoring the congruent orders? The answer to this question is NO. Although there lack large-scale typological studies of adjective orders in world's languages, as shown in simple Chinese expressions, the adjectives of textural patterns usually appear in front of the color or shape adjectives. Although in many English phrases, the shape adjectives should appear in front of the color adjectives, most people have a relatively free order between the two. Considering these, apart from the perceptual saliency hierarchy, the structural configuration of language is also subject to other constraints. One candidate of such constraints comes from the socio-cultural environment of language. As shown in typological studies of structural diversity in world's languages (Haspelmath, 2007; Evans and Levinson, 2009; Dunn et al., 2011), cultural histories of speakers and contact histories between different languages could induce different types of structure.

In addition, our experiment showed that despite the fact that the participants exhibited biases towards the congruent orders, after a small number of learning rounds, they could largely grasp both the biased and unbiased orders, reaching high (over 0.8) accuracy and short reaction time. Following the dynamics in the three experimental phases, we can reasonably expect that given more rounds of learning the participants would learn each artificial language equally well, no matter whether its order was congruent or incongruent. This suggests that the structures distinct from the biased ones can be equally acquired by speakers. This makes sure that other types of structures, once induced due to other constraints, can also be transmitted across generations of learners.

The above discussion reveals a complicated relation between language and perception.

On the one hand, during cultural transmission of language across multiple generations of learners, individuals' perceptual constraints could favor some structures congruent with the perceptual constraints, thus causing a bias towards those structures. This has been demonstrated in many experimental and simulation studies. For example, some experiments have

shown that the dominant word orders in world's languages are also easier to learn (Culbertson et al., 2012; Fedzechkina et al., 2012; Culbertson and Adger, 2014). Some simulations have also revealed that cultural transmission could amplify small biases for certain structure and make it prevalent in communal languages of later generations (Griffiths and Kalish, 2007; Kirby et al., 2008; Smith, 2011).

On the other hand, other factors, such as different socio-cultural histories could induce distinct language structures. Some modeling studies have demonstrated that socio-cultural interactions could trigger a variety of structural forms, which can be equally acquired and transmitted by generations of language learners (Steels and Belpaeme, 2005; Steels, 2011, 2012). More importantly, as shown in our study, if different structures are more or less functionally equivalent, they can be acquired equally well by speakers, given sufficient rounds of learning. This may diminish the bias for certain structures to a certain extent, and lead to diversity in structural configuration of language.

These two aspects suggest that the actual structures in different languages have arisen as a compromise between both the individual perceptual constraints and the socio-cultural factors (Christiansen and Chater, 2008; Liu, 2014). Such compromise leads to a biased distribution of languages predominantly in certain structures. The mutual influence of individual and socio-cultural factors has been illustrated in some simulation studies of word order bias (Gong et al., 2009) and color naming patterns (Baronchelli et al., 2012).

Our experiment, as an individual learning experiment, could not fully demonstrate the mutual influence of individual and socio-cultural aspects. Nonetheless, it confirmed the influence of individual perceptual constraint (i.e., the perceptual saliency hierarchy) on learning congruent and incongruent orders. It also revealed that given more trials the learning of incongruent orders could reach a similar level to the learning of congruent orders. This suggested that the structures induced by other factors, even though conflicting to the structures favored by perceptual constraints, could still be acquired and transmitted. Both of these findings can shed light on the relation between individual learning and cultural transmission, and contribute to the discussion of the causal factors for the structural diversity in languages (Longobardi and Guardiano, 2009).

Finally, some aspects of this study can be extended in future work. For example, we may recruit pre-language or language-learning children to further diminish the influence of individuals'

prior linguistic knowledge (Saffran et al., 2007; Folia et al., 2010). Compared with visual presentation, auditory presentation better resembles language exchange in everyday life. It may reduce the effect of orthography in language learning, though not eliminating it altogether (Cuskley et al., 2015). In auditory presentation, factors such as memory (Morrison et al., 2014), stress or prosody may also modulate the biases for certain structures.

## ETHICS STATEMENT

The experimental protocol was approved by the College Research Ethics Committee of University of Hong Kong. The methods were carried out in accordance with the approved guidelines from the College Research Ethics Committee. Written informed consents were obtained from all participants.

## AUTHOR CONTRIBUTIONS

TG and LS designed the research, YL carried out the study. TG and YL analyzed the results. TG, LS, and YL wrote the paper.

## ACKNOWLEDGMENTS

TG is supported by the US NIH Grant (HD-071988). The study has been supported in part by the MOE Project of the Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies. The preliminary results were first presented in the 10th International Conference on the Evolution of Language (EVOLANG 10) in 2014. We thank Luke Yiu from University of Hong Kong and David Kush from Haskins Laboratories for their comments on this work.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01952/full#supplementary-material>

## REFERENCES

- Baronchelli, A., Gong, T., Puglisi, A., and Loreto, V. (2012). Modelling the emergence of universality in color naming patterns. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2403–2407. doi: 10.1073/pnas.0908533107
- Braine, M. D. S. (1963). On learning the grammatical order of words. *Psychol. Rev.* 70, 323–348. doi: 10.1037/h0047696
- Brighton, H. (2002). Compositional syntax from cultural transmission. *Artif. Life* 8, 25–54. doi: 10.1162/106454602753694756
- Brooks, P. J., Kempe, V., and Donachie, A. (2011). Second language learning benefits from similarity in word endings: evidence from Russian. *Lang. Learn.* 61, 1142–1172. doi: 10.1111/j.1467-9922.2011.00665.x
- Callaghan, T. C. (1990). "Interference and dominance in texture segregation," in *Visual Search*, ed. D. Brogan (New York, NY: Taylor & Francis), 81–87.
- Carter, R., McCarthy, M., Mark, G., and O'Keefe, A. (2011). *English Grammar Today: An A-Z of Spoken and Written Grammar*. London: Cambridge University Press.
- Chater, N., Reali, F., and Christiansen, M. H. (2009). Restrictions on biological adaptation in language evolution. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1015–1020. doi: 10.1073/pnas.0807191106
- Christiansen, M. H., and Chater, N. (2008). Language as shaped by the brain. *Behav. Brain Sci.* 31, 489–509. doi: 10.1017/S0140525X08004998
- Christiansen, M. H., and Chater, N. (2015). The language faculty that wasn't: a usage-based account of natural language recursion. *Front. Psychol.* 6:1182. doi: 10.3389/fpsyg.2015.01182

- Christiansen, M. H., and Ellefson, M. R. (2002). "Linguistic adaptation without linguistic constraints: the role of sequential learning in language evolution," in *The Transition to Language*, ed. A. Wray (Oxford: Oxford University Press), 335–358.
- Christiansen, M. H., and Kirby, S. (2003). Language evolution: consensus and controversies. *Trends Cogn. Sci.* 7, 300–307. doi: 10.1016/S1364-6613(03)00136-0
- Cinque, G. (1999). *Adverbs and Functional Heads: A Cross-linguistic Perspective*. London: Oxford University Press.
- Cinque, G. (2010). *The Syntax of Adjectives: A Comparative Study*. Cambridge, MA: MIT Press.
- Cinque, G., and Rizzi, L. (2008). The cartography of syntactic structures. *Stud. Ling.* 2, 42–58. doi: 10.1007/s10936-014-9330-6
- Claidière, N., Smith, K., Kirby, S., and Fagot, J. (2014). Cultural evolution of a systematically structured behavior in a non-human primate. *Proc. R. Soc. B* 281: 20141541. doi: 10.1098/rspb.2014.1541
- Conway, C. M., and Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends Cogn. Sci.* 5, 539–546. doi: 10.1016/S1364-6613(00)01800-3
- Cornish, H. (2010). Investigating how cultural transmission leads to the appearance of design without a designer in human communication systems. *Interact. Stud.* 11, 112–137. doi: 10.1075/is.11.1.02cor
- Culbertson, J., and Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5842–5847. doi: 10.1073/pnas.1320525111
- Culbertson, J., Smolensky, P., and Legendre, G. (2012). Learning biases predict a word order universal. *Cognition* 122, 306–329. doi: 10.1016/j.cognition.2011.10.017
- Cuskley, C., Simner, J., and Kirby, S. (2015). Phonological and orthographic influences in the bouba-kiki effect. *Psychol. Res.* doi: 10.1007/s00426 [Epub ahead of print].
- De Boer, D., and Verhoef, T. (2012). Language dynamics in structured form and meaning spaces. *Adv. Complex Syst.* 15:1150021. doi: 10.1142/S0219525911500214
- Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Lang. Ling. Comp.* 6, 654–672. doi: 10.1002/lnc3.361
- Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., and Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends Cogn. Sci.* 19, 603–615. doi: 10.1016/j.tics.2015.07.013
- Dominy, N. J., and Lucas, P. W. (2001). Ecological importance of trichromatic vision to primates. *Nature* 410, 363–366. doi: 10.1038/35066567
- Dunn, M., Green, S. J., Levinson, S. C., and Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473, 79–82. doi: 10.1038/nature09923
- Elman, J. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1207/s15516709cog1402\_1
- Esper, E. A. (1925). A technique for the experimental investigation of associative interference in artificial linguistic material. *Lang. Mono.* 1, 1–47.
- Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., and Wong, P. C. M. (2015). The relationship between artificial and second language learning. *Cogn. Sci.* 40, 822–847. doi: 10.1111/cogs.12257
- Evans, N., and Levinson, S. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32, 429–492. doi: 10.1017/S0140525X0999094X
- Fedzechkina, M., Jaeger, T. F., and Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17897–17902. doi: 10.1073/pnas.1215776109
- Ferrer-i-Cancho, R. (2004). Euclidean distance between syntactically linked words. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 70(5 Pt 2):056135. doi: 10.1103/PhysRevE.70.056135
- Folia, V., Uddén, J., de Vries, M., Forkstam, C., and Petersson, K. M. (2010). Artificial language learning in adults and children. *Lang. Learn.* 60(Suppl. 2), 188–220. doi: 10.1111/j.1467-9922.2010.00606.x
- Friederici, A. D., Steinhauer, K., and Pfeifer, E. (2002). Brain signatures of artificial language processing: evidence challenging the "critical period" hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 529–534. doi: 10.1073/pnas.012611199
- Futrell, R., Mahowald, K., and Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10336–10341. doi: 10.1073/pnas.1502134112
- Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cogn. Sci.* 29, 737–767. doi: 10.1207/s15516709cog0000\_34
- Galantucci, B., and Garrod, S. (2010). Experimental semiotics: A new approach for studying the emergence and the evolution of human communication. *Interact. Stud.* 11, 1–13. doi: 10.1075/is.11.1.01gal
- Galantucci, B., and Garrod, S. (2011). Experimental semiotics: a review. *Front. Hum. Neurosci.* 5:11. doi: 10.3389/fnhum.2011.00011
- Gentner, T. Q., and Goldin-Meadow, S. (eds) (2003). *Language in Mind: Advances in the Study of Language and Thought*. Cambridge, MA: MIT Press.
- Gong, T., Minett, J. W., and Wang, W. S.-Y. (2009). A simulation study on word order bias. *Interact. Stud.* 10, 51–75. doi: 10.1075/is.10.1.04gon
- Gong, T., Shuai, L., and Wu, Y. (2013). Multidisciplinary approaches in evolutionary linguistics. *Lang. Sci.* 37, 1–13. doi: 10.1111/brv.12160
- Gong, T., Shuai, L., and Zhang, M. (2014). Modelling language evolution: examples and predictions. *Phys. Life Rev.* 11, 280–302. doi: 10.1016/j.plrev.2013.11.009
- Griffiths, T. L., and Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cogn. Sci.* 31, 441–480. doi: 10.1080/15326900701326576
- Gómez, R. L., and Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends Cogn. Sci.* 4, 178–186. doi: 10.1016/S1364-6613(00)01467-4
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English (part II). *J. Ling.* 3, 199–244. doi: 10.1017/S0022226700016613
- Haspelmath, M. (2007). Pre-established categories don't exist: consequences for language description and typology. *Ling. Typol.* 11, 119–132.
- Healey, C. G. (2000). Building a perceptual visualization architecture. *Behav. Inf. Technol.* 19, 349–366. doi: 10.1080/014492900750000054
- Healey, C. G., and Enns, J. T. (1999). Large datasets at a glance: Combining textures and colors in scientific visualization. *IEEE Trans. Vis. Comput. Graph.* 5, 145–167. doi: 10.1109/2945.773807
- Healey, C. G., and Enns, J. T. (2012). Attention and visual memory in visualization and computer graphics. *IEEE Trans. Vis. Comput. Graph.* 18, 1170–1188. doi: 10.1109/TVCG.2011.127
- Heine, B. (2001). *The Cognitive Foundations of Grammar*. Oxford: Oxford University Press.
- Heine, B., and Kuteva, T. (2008). *The Genesis of Grammar: A Reconstruction*. Oxford: Oxford University Press.
- Hurford, J. (2007). *The Origins of Meaning*. Oxford: Oxford University Press.
- Hurford, J. (2012). *The Origins of Grammar*. Oxford: Oxford University Press.
- Kirby, S. (1999). *Function, Selection and Innateness: The Emergence of Language Universals*. Oxford: Oxford University Press.
- Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10681–10686. doi: 10.1073/pnas.0707835105
- Kirby, S., Dowman, M., and Griffiths, T. (2006). Innateness and culture in the evolution of language. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5241–5245. doi: 10.1073/pnas.0608222104
- Lambrecht, K. (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- Liu, H. T. (2010). Dependency distance as a means of word-order typology: a method based on dependency treebanks. *Lingua* 120, 1567–1578. doi: 10.1016/j.lingua.2009.10.001
- Liu, H. T. (2014). Language is more a human-driven system than a semiotic system: comment on "Modelling language evolution: examples and predictions" by Tao Gong, Lan Shuai, Menghan Zhang. *Phys. Life Rev.* 11, 309–310. doi: 10.1016/j.plrev.2013.12.008
- Longobardi, G., and Gardiano, C. (2009). Evidence for syntax as a signal of historical relatedness. *Lingua* 119, 1679–1706. doi: 10.1016/j.lingua.2008.09.012
- Lupyan, G., and Christiansen, M. H. (2002). "Case, word order, and language learnability: Insights from connectionist modeling," in *Proceedings 24th Annual Cognitive Science Society*, eds W. D. Gray and C. Schunn (Mahwah, NJ: Lawrence Erlbaum Associates), 596–601.
- Mathesius, V. (1975). *A Functional Analysis of Present Day English on a General Linguistic Basis*, ed. J. Vachek. The Hague: Mouton.
- Melin, A. D., Hiramatsu, C., Fedigan, L. M., Schaffner, C. M., Aureli, F., and Kawamura, S. (2012). "Ploymorphism and adaptation of primate color vision,"



- in *Evolutionary Biology: Mechanisms and Trends*, ed. P. Pontarotti (London: Springer), 225–241.
- Mesoudi, A. (2011). *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*. Chicago, IL: University of Chicago Press.
- Mesoudi, A., and Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philos. Trans. R. Soc. B* 363, 3489–3501. doi: 10.1098/rstb.2008.0129
- Misyak, J. B., and Christiansen, M. H. (2012). Statistical learning and language: an individual differences study. *Lang. Learn.* 62, 302–331. doi: 10.1111/j.1467-9922.2010.00626.x
- Misyak, J. B., Christiansen, M. H., and Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Front. Psychol.* 1:31. doi: 10.3389/fpsyg.2010.00031
- Moeser, D. S., and Bregman, A. S. (1972). The role of reference in the acquisition of a miniature artificial language. *J. Verbal Learn. Verbal Behav.* 11, 759–769. doi: 10.1016/S0022-5371(72)80010-0
- Morgan, J. L., and Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *J. Verbal Learn. Verbal Behav.* 20, 67–85. doi: 10.1016/S0022-5371(81)90312-1
- Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K., Carpenter, H., and Wong, P. C. M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism* 17, 56–72. doi: 10.1017/S1366728912000715
- Morrison, A. B., Conway, A. R. A., and Chien, J. M. (2014). Primacy and recency effects as indices of the focus of attention. *Front. Hum. Neurosci.* 8:6. doi: 10.3389/fnhum.2014.00006
- Onnis, L. (2012). “Artificial language learning,” in *Routledge Encyclopedia of Second Language Acquisition*, ed. P. Robinson (New York, NY: Routledge), 5–39.
- Osborne, J. W., and Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Pract. Assess. Res. Eval.* 9, 1–12.
- Pemiss, P., and Vigliocco, G. (2014). The bridge of iconicity: from a world experience to the experience of language. *Philos. Trans. R. Soc. B* 369:20130300. doi: 10.1098/rstb.2013.0300
- Petersson, K. M., Folia, V., and Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain Lang.* 120, 83–95. doi: 10.1016/j.bandl.2010.08.003
- Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychol. Bull.* 133, 227–244. doi: 10.1037/0033-2909.133.2.227
- Pullum, G. K., and Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *Ling Rev* 19, 9–50.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reali, F., and Christiansen, M. H. (2009). Sequential learning and the interaction between biological and linguistic adaptation in language evolution. *Interact. Stud.* 10, 5–30. doi: 10.1075/is.10.1.02rea
- Reber, A. S. (1967). Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* 6, 855–863. doi: 10.1016/S0022-5371(67)80149-X
- Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge*. New York, NY: Oxford University Press.
- Richerson, P. J., and Christiansen, M. H. (eds) (2013). *Cultural Evolution: Society, Technology, Language, and Religion*. Cambridge, MA: MIT Press.
- Robinson, P. (2010). Implicit artificial grammar and incidental natural second language learning: how comparable are they? *Lang. Learn.* 60(Suppl. 2), 245–263. doi: 10.1111/j.1467-9922.2010.00608.x
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926
- Saffran, J. R., Pollak, S. D., Seibel, R. L., and Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition* 105, 669–680. doi: 10.1016/j.cognition.2006.11.004
- Scott-Phillips, T. C., Kirby, S., and Ritchie, G. R. S. (2009). Signaling signalhood and the emergence of communication. *Cognition* 113, 226–233. doi: 10.1016/j.cognition.2009.08.009
- Simmer, J., Cuskley, C., and Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception* 39, 553–569. doi: 10.1068/p6591
- Smith, K. (2011). Learning bias, cultural evolution of language, and the biological evolution of the language faculty. *Hum. Biol.* 83, 261–278. doi: 10.3378/027.083.0207
- Smith, K., Brighton, H., and Kirby, S. (2003). Complex systems in language evolution: the cultural emergence of compositional structure. *Adv. Complex Syst.* 6, 537–558. doi: 10.1371/journal.pone.0037744
- Steels, L. (2011). Modelling the cultural evolution of language. *Phys. Life Rev.* 8, 339–356. doi: 10.1016/j.plrev.2011.10.014
- Steels, L. (2012). *Experiments in Cultural Language Evolution*. Amsterdam: John Benjamins.
- Steels, L., and Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: a case study for color. *Behav. Brain Sci.* 28, 469–488. doi: 10.1017/S0140525X05000087
- Tabullo, Á, Arismendi, M., Wainelboime, A., Primero, G., Vernis, S., Segura, E., et al. (2012). On the learnability of frequent and infrequent word orders: an artificial language learning study. *Q. J. Exp. Psychol.* 65, 1848–1863. doi: 10.1080/17470218.2012.677848
- Tamariz, M., and Kirby, S. (2015). Culture: copying, compression, and conventionality. *Cogn. Sci.* 39, 171–183. doi: 10.1111/cogs.12144
- Taylor, J. S. H., Plunkett, K., and Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: an artificial orthography paradigm. *J. Exp. Psychol.* 37, 60–76. doi: 10.1037/a0020126
- Tily, H., Frank, M. C., and Jaeger, T. G. (2011). The learnability of constructed languages reflects typological patterns. *Cogn. Sci. Soc.* 33, 1364–1369.
- Tomlin, R. S. (1986). *Basic Word Order: Functional Principles*. London: Croom Helm.
- Treisman, A. (1985). Preattentive processing in vision. *Comput. Vis. Graph Image Process.* 31, 156–177. doi: 10.1016/S0734-189X(85)80004-9
- Yip, P.-C., and Rimmington, D. (2004). *Chinese: A Comprehensive Grammar*. London: Routledge.
- Zhu, Q. (2005). *Modern Chinese Practical Grammar Analysis*. Beijing: Tsinghua University Press.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Gong, Lam and Shuai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Reversing the Approach to Null Subjects: A Perspective from Language Acquisition

Maia Duguine\*

Linguistics and Basque Studies, University of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain

## OPEN ACCESS

### Edited by:

Ángel J. Gallego,  
Autonomous University of Barcelona,  
Spain

### Reviewed by:

Anna Cardinaletti,  
Ca' Foscari University of Venice, Italy  
Elisabet Pladevall-Ballester,  
Universitat Autònoma de Barcelona,  
Spain

Olga Fernández-Soriano,  
Universidad Autónoma de Madrid,  
Spain

### \*Correspondence:

Maia Duguine  
maia.duguine@ehu.eus

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 11 October 2016

**Accepted:** 04 January 2017

**Published:** 14 February 2017

### Citation:

Duguine M (2017) Reversing the  
Approach to Null Subjects: A  
Perspective from Language  
Acquisition. *Front. Psychol.* 8:27.  
doi: 10.3389/fpsyg.2017.00027

This paper proposes a new model for null subjects, and focuses on its implications for language development. The literature on pro-drop generally considers that not allowing null subjects is, informally speaking, the “default” option in natural languages, and appeals to particular morphosyntactic mechanisms in order to account for those languages in which the subject can be omitted. Shifting the perspective, the *inverse approach* postulates that pro-drop is (almost) a default grammatical setting, and that non-pro-drop results from the intervention of independent factors that block pro-drop in the derivation. The paper explores the consequences of the inverse approach in the domain of language acquisition, arguing that this model allows to account for a number of properties of child languages. It opens an avenue of research worth exploring, one that could give new solutions to old problems.

**Keywords:** pro-drop, null subjects, language acquisition, case, language variation

## 1. INTRODUCTION

Under specific circumstances, sentences can have subjects which, even if unpronounced, are syntactically projected (see recently Cai et al., 2014). Pro-drop, or the possibility to omit the subject of a finite construction is a phenomenon which, in theoretical linguistics, is generally studied from a comparative perspective. The central question is why certain languages allow null subjects while others do not<sup>1</sup>. There is thus an opposition between pro-drop languages vs. non-pro-drop languages. But these studies have very often led to an (implicitly) asymmetrical characterization of the two options. That is, in a sense, it is considered that non-pro-drop is the default option in natural languages, and that the pro-drop option has to be motivated (that is, it has to be explained by appealing to a particular grammatical mechanism)<sup>2</sup>. Indeed, languages are taken to need a special grammatical feature in order to allow pro-drop, such as for instance, a pronominal Agr (cf. Rizzi, 1982; Alexiadou and Anagnostopoulou, 1998), a [D] feature in T (cf. Holmberg, 2010; Roberts, 2010a), special Case-assigners (Rizzi, 1986), or uniform agreement (Jaeggli and Safir, 1989).

A further asymmetrical characterization emerges from the work on individual or groups of pro-drop languages. It is often assumed that natural languages offer multiple ways of licensing null subjects, and thus, different types of pro-drop languages are also postulated. For instance,

<sup>1</sup>For reasons of space, and in order to be able to focus on the central aims of the paper, I will leave null objects aside (see footnote 16). Therefore, in this paper I use the terms *pro-drop*, *subject-drop* and *null subject* interchangeably.

<sup>2</sup>Note that this statement is not equivalent to assuming that there is a default negative setting for the pro-drop parameter: it just aims at making explicit how researchers approach the issue of null arguments. See also footnote 18.

Italian-type languages and Chinese-type languages are typically distinguished, as allowing pro-drop vs. topic-drop (cf. Huang, 1984), or as licensing *pro* vs. argument ellipsis (cf. Saito, 2007; Roberts, 2010a). Again, this implies that conceptually, we are considering non-pro-drop as being the default option, and the fact that some languages allow null subjects is taken to derive from additional features these languages have.

But we could also flip this idea, and consider that in a sense, it is null subjects being possible that constitutes the default option, and that what has to be explained is non-pro-drop, in terms of a set of cases in which subject-drop is made impossible in the derivation of the sentence. Under this view, pro-drop is the same phenomenon in Italian and Chinese; what requires an explanation is the impossibility to drop subjects in English or French, for instance.

Furthermore, viewing pro-drop as an operation that can be blocked allows us to appeal to different conditioning factors in different cases. The same way movement can be blocked by an island, or an intervener, or the landing site being already filled, there are potentially multiple ways in which pro-drop will be blocked. This has the potential to explain the variety of cases where null subjects are not allowed, across different languages (in non-pro-drop languages, but also in pro-drop languages; see below). I will call this view the *inverse approach to pro-drop* (IA).

Now, null subjects constitute one of the best-studied topics in language acquisition, which has enlightened many aspects of the discussion on the logical problem of language acquisition, the nature of variation, parametric theory, etc. (see Hyams, 2011; Hyams et al., 2015, for recent overviews of the literature). Does the IA and the shifted view it proposes have something to bring to this field? And to what extent do the developmental data, observations and generalizations that have been collected and discovered over the years conform to the model that is suggested by the IA?

Taking the IA as a reference, the goal of this paper is to open a new perspective on the topic of null subjects in the area of acquisition, and as a first step, to explore the extent to which what we know about the acquisition of the pro-drop property makes sense under the IA.

Section 2 introduces the basic components of an account of null subjects that formalizes the fundamental ideas of the IA, and briefly presents typological and empirical evidence supporting this view. Section 3 explores some of the consequences of the shift to IA for the domain of language acquisition, on whether the standard observations on the stages of acquisition of pro-drop can be accounted for straightforwardly. Section 5 gives the conclusions.

## 2. REVERSING THE PERSPECTIVE

This section introduces the basic features of the *inverse approach to pro-drop* (IA). It does not propose a full-fledged analysis of pro-drop (see Duguine, 2013, 2014 for a more elaborated proposal). Rather, it sketches a possible account that would formalize the basic ideas of the IA that were introduced above. It also discusses evidence that supports these ideas.

### 2.1. Pro-Drop and Non-Pro-Drop under the Inverse Approach

By characterizing pro-drop as the “default” option for a language *L*, we do not necessarily have to assume that pro-drop is totally free and not subject to any syntactic condition. Instead, the claim is that all natural languages satisfy the very basic syntactic condition for allowing it, and that if a language happens not to allow null subjects, this is a fact that has to be explained.

Observe the following examples from Spanish, a pro-drop language. Whereas the DP *todos los días* “all days” can alternate with a null expression in (1B), it cannot in (2B) (the null subject is represented with “[e]”):

- (1) A. Todos los días son una fiesta. *Spanish*  
 all the days are a party  
 Every day is a party.  
 B. No, [e] no son una fiesta.  
 NEG NEG are a party  
 No, they are not a party.
- (2) A. Yo salgo de fiesta todos los días.  
 I go.out.1sg of party every the days  
 I go out to party every day.  
 B. Yo no salgo de fiesta (\*[e]).  
 I NEG go.out.1sg of party  
 I don't go out to party.

(2B) cannot be interpreted as “I don't go out to party every day.” This shows that pro-drop is subject to a syntactic constraint. But what is this constraint? It is fair to say that the basic mechanism that makes pro-drop possible is the one behind the argument-adjunct distinction: arguments can drop, adjuncts cannot. Let us assume that structural Case—in particular, nominative-assignment in the case of subjects (and, potentially, ergative)—is this mechanism (see a.o. Chomsky, 1982; Raposo, 1986; Rizzi, 1986; Jaeggli and Safir, 1989; Platzack and Holmberg, 1989, where Case is defined as the basis of the “licensing” condition for *pro*)<sup>3</sup>. Assuming that Case operations hold in all natural languages, this makes all languages potential pro-drop languages. In other words, by default, any language will allow null arguments. In particular, given the (arguably universal) Case-assigning properties of finite T, this analysis accounts for the availability of null subjects across languages. What has to be accounted for are thus those languages that do not allow null arguments (or more specifically, null subjects), i.e., non-pro-drop languages.

The idea, under the IA, is that in these languages, even if the Case condition is satisfied—and thus pro-drop is in principle available—, independent factors come into play which block pro-drop. This idea can be illustrated with cases in which null subjects are impossible in pro-drop languages. It is for instance well-known that there are no focused null subjects (cf. Cardinaletti and Starke, 1999). In fact, focused subjects are always overt (cf. Larson and Luján, 1989), as illustrated in the Spanish question-answer pair in (3) (capital letters indicate focusing):

<sup>3</sup>For a discussion of evidence in favor of Structural Case as the condition on pro-drop, see Duguine (2013).

- (3) A. Juan ha leído Guerra y paz. *Spanish*  
 Juan have.3sg read war and peace  
 Juan has read War and peace.  
 B. No: lo he leído YO/\*[e].  
 no CL have.1sg read I  
 No: I read it.

That is, focus has a blocking effect on pro-drop, even in contexts in which the subject satisfies the conditions for being null (i.e., it is assigned nominative Case).

In line with this observation, the hypothesis I will put forth under the perspective of the IA is that non-pro-drop languages are languages in which there is always, in the derivation, something that blocks pro-drop. What could it be? There is a long-standing hypothesis in the literature on pro-drop, that connects the “richness” of subject-verb agreement morphology with the availability of null subjects. Indeed, pro-drop languages that have agreement morphology—such as Spanish or Italian—tend to have “rich” inflectional systems (with different forms for different person-number affixes), whereas non-pro-drop languages such as English or German tend to have many syncretic forms, i.e., “poor” agreement. This is the so-called “Taraldsen’s generalization” (Taraldsen, 1980; Jaeggli and Safir, 1989)<sup>4</sup>. Many analyses have built on this generalization, defending that “rich” agreement is what makes null subjects possible (cf. Barbosa, 1995; Alexiadou and Anagnostopoulou, 1998; Speas, 2006). Following the logic of the IA, I would like to suggest here that we reverse the perspective, and postulate that in fact, “rich” agreement is not a condition on pro-drop; instead, it is “poor” agreement that blocks pro-drop.

This hypothesis can be formalized using Frampton’s (2002) and Müller’s (2006, 2008) characterization of poor inflection as impoverished inflection. Under the Distributed Morphology approach, impoverishment is an operation that deletes morphosyntactic features on abstract morphemes in certain specific contexts (cf. a.o. Bonet, 1991; Halle, 1997; Harley and Noyer, 1999). A morpheme which undergoes this operation ends up with a set of features less specified than it was before the operation took place. Frampton (2002) and Müller (2006, 2008) propose that in languages such as German, there are certain impoverishment operations which systematically delete (valued)  $\varphi$ -features on T, leading to the feature-specification of different morphemes being identical and thus to have the same phonological realization.

Crucially, developing the intuition that poor agreement is actually impoverished agreement, Müller (2006) makes the following suggestion: impoverishment actually bleeds pro-drop. He explains non-pro-drop in the following terms<sup>5</sup>:

<sup>4</sup>“Richness” has proven to be a difficult notion to define (cf. attempts in a.o. Jaeggli and Safir, 1989; Rohrbacher, 1999; Müller, 2006). Nonetheless, the generalization holds, which suggests that at least at an abstract level, “richness” is relevant for pro-drop (cf. also Roberts, 1993; Platzack, 1994; Vainikka and Levy, 1999).

<sup>5</sup>Müller (2006, 2008)’s analysis relies on the assumption that Morphological Structure comes before syntax proper, so that  $\varphi$ -Agree—a syntactic operation—can be affected by the output of impoverishment—a morphological operation. See Duguine (2013) for an alternative analysis that maintains a standard architecture of the grammar, with a post-syntactic morphological component.

- (4) *Pro* generalization (Müller, 2006)  
 An argumental *pro* DP cannot undergo Agree with a functional head  $\alpha$  if  $\alpha$  has been subjected (perhaps vacuously) to  $\varphi$ -feature neutralizing impoverishment in the numeration.

That is, like any subject DP, *pro* enters a  $\varphi$ -Agree relation with T. But in contrast to DPs, it cannot enter such a relation if T has been impoverished. This would account for why subjects are necessarily overt in languages in which  $\varphi$ -features on T undergo impoverishment (see also Roberts, 2010b; Duguine, 2013).

Summarizing, the IA postulates that pro-drop “comes for free” in natural languages, and that non-pro-drop is what must be accounted for. As a way of formalizing this idea, on the one hand, I have proposed that structural Case-assignment is what makes null arguments available. Under the assumption that Case relations are a pervasive feature of languages, this implies that all languages are, in principle, potential pro-drop languages. It also accounts for pro-drop in all types of languages in which arguments can be null. In particular, it invites to a unified analysis of Italian-like and Japanese-like pro-drop languages (see Duguine, 2014 for arguments in favor of this unification). On the other hand, in order to account for languages that do not allow null subjects, I have appealed to the analysis proposed by Müller (2006, 2008), whereby non-pro-drop results from independent factors: impoverished T cannot combine with a null subject.

Note finally that the explanation of the non-pro-drop option in terms of impoverishment is just one example of how pro-drop can be blocked. The case of focus, discussed above, shows that there can in principle be many different ways in which different factors affect pro-drop. For instance, it has been proposed that the fact that English is not a null subject language results from T requiring an overt specifier (cf. Holmberg, 2010). If this analysis is on the right track, then it could be that in this case it is not impoverished inflection that blocks pro-drop, but rather this overt condition on Spec,TP. The IA thus leads to a potentially multimodular and multifactorial characterization of the (non-) pro-drop phenomenon.

## 2.2. Typological and Empirical Evidence

The picture offered by the IA is rather unusual: it implies that pro-drop is a universal phenomenon, available in principle across all languages, with exceptions that will have to be accounted for on independent grounds. Nonetheless, as expected under this view, the availability of null arguments seems to be the unmarked option cross-linguistically.

Null arguments are licensed in the majority of the languages of the world. The broadest survey of pro-drop is probably the one by Dryer (2013), in the *World Atlas of Linguistics Structures*, which focuses on the way in which subjects are—or can be—expressed. Spanish-type languages and Japanese-type languages (i.e., pro-drop languages with and without agreement) represent 70% of the sample of languages analyzed by Dryer (2013) (498 out of 711). On the other hand, languages in which “pronominal subjects are expressed by pronouns in subject position that are normally if not obligatorily present” (English, German, French, Icelandic, etc.) represent 11.5% of the total

number of languages<sup>6</sup>. 70% constitutes a very large majority, and the quantitative difference between pro-drop languages and non-pro-drop languages is significant<sup>7</sup>.

Also, the IA characterizes non-pro-drop as a property of derivations, and not as a defining property of languages. Precisely, there is a sense in which non-pro-drop languages are not fully non-pro-drop, given that there are cases, contexts or varieties in which they allow null subjects. For instance, (i) subjects of imperatives tend to be null (*cf.* Bennis, 2006 on Dutch), (ii) null subjects of finite matrix and embedded clauses are observed in certain varieties of English, such as diary British English (Haegeman and Ihsane, 2001) or Colloquial Singapore English (Sato, 2011; Sato and Kim, 2012), (iii) null subjects are also licensed in certain varieties of French—one of the few non-pro-drop Romance languages (*cf.* Roberge, 1990; Zribi-Hertz, 1994, as well as Roberts, 2010b for a critical review of the data), and (iv) Rosenkvist (2009) emphasizes that, even if null subjects are licensed in none of the modern Germanic standard languages, they are in many modern vernaculars (Zürich German, Schwabian, Bavarian, Lower Bavarian, Frisian, Övdalian and Yiddish).

In sum, the dichotomy between “pro-drop languages” vs. “non-pro-drop languages” has been largely overestimated in the literature. Indeed, the cross-linguistic data suggest that allowing null subjects is the default option for languages, and that we are not dealing with a phenomenon deeply rooted in the nature of languages, but rather the result of the conspiracy of unrelated factors affecting the derivation, as implied by the IA.

### 3. A NEW PERSPECTIVE ON PRO-DROP IN ACQUISITION

Given the approach outlined in the previous section, the obvious question from a developmental perspective is to ask whether it can help us reach an explanation of the acquisition process. Indeed, the IA's shift regarding the question of null arguments does not have consequences for the theory of syntax only; it also affects how acquisition of the pro-drop property is expected to take place. This section explores the question of whether the IA makes sense from the point of view of language development. To that end, it briefly reviews a set of basic facts that have been established in the literature on the acquisition of (non-)pro-drop, and attempts to evaluate whether they correspond to what we could expect under the IA.

#### 3.1. Early Subject Omission in Pro-Drop Languages

Speakers of pro-drop languages show target-like behavior from very early on (see Valian, 1990; Guasti, 1993/1994 on Italian, Valian and Eisenberg, 1996 on European Portuguese,

<sup>6</sup>The three other groups of languages—which constitute 18.4% of the sample—are not easy to classify directly as either pro-drop or non-pro-drop, but some of them, such as Warlpiri (Legate, 2006), Finnish, Hebrew (Vainikka and Levy, 1999) or Irish (*cf.* McCloskey and Hale, 1984) are known to allow null subjects. The actual proportion of null subject languages is thus larger than 70%.

<sup>7</sup>In the sample of 104 languages studied in Gilligan (1987), 93% are classified as null subject languages.

Wang et al., 1992 on Chinese, Kim, 1997 on Korean among others).

Under the IA, pro-drop is a default or given property of languages<sup>8</sup>. Therefore, the observation that children acquiring a pro-drop language show a target-like behavior is consistent with what we could expect given the IA. It is nonetheless important to note that this is not a prediction. The syntax of pro-drop is, logically, dependent on the syntax of subjects, and in particular, as proposed in Section 2, on the syntax of (structural) Case. Therefore, a child will not be expected to drop subjects until she has acquired the syntax of subjects and their Case properties (on the role of Case in the syntax and acquisition of pro-drop, see also Pierce, 1992). Consequently, what the IA predicts is that the possibility to drop the subject will follow the acquisition of the syntax of subjects. In other words, given the early acquisition of null subjects, we expect an early acquisition of the syntax of subject's Case in pro-drop languages. Precisely, acquisition of pro-drop languages seems to be characterized by an early knowledge of the syntax of subjects. For instance, in pro-drop languages, children start producing inflected verbal forms (with virtually no errors in person-agreement) and target-like subject placements very early on (*cf.* among others Guasti, 1993/1994 on Italian, Bel, 2003 on Spanish, and Barreña, 1995; Ezeizabarrena, 2002 on Basque).

#### 3.2. Null Subjects in Early Non-Pro-Drop Languages

It is well known that early non-pro-drop languages such as English, Dutch or French allow null subjects (*cf.* Hyams, 1986). As we just saw, under the IA, given the “default” nature of pro-drop, setting the syntax of subjects is sufficient for allowing null subjects. As above, the prediction is therefore that the syntax of subjects, and in particular Case-assignment is in place from very early on in non-pro-drop languages, too.

Here, too, the prediction seems to be on the right track. Schütze and Wexler (1996) show that in early English virtually all (pronominal) subjects of finite verbs are nominative, unlike the subjects of non-finite verbs, which are often accusative (see below on root infinitives). Since in English accusative—but not nominative—is the default case (that is, DPs surface with accusative marking when they are not assigned Case; *cf.* Schütze, 2001), we can conclude with Schütze and Wexler (1996) that the fact that subjects in finite contexts are virtually always nominative shows that the syntax of nominative Case is already in place for those speakers.

#### 3.3. Later Setting of the Non-Pro-Drop Option

The third point is closely related to the preceding two. The observation is that whereas speakers of null subject languages seem to have a very early acquisition of the pro-drop property of their target language (i.e., what Hoekstra and Hyams, 1998 call

<sup>8</sup>This consequence of the IA converges with the early parameter missetting approach in Hyams (1986), Jaeggli and Hyams (1988), and Hyams (1991), which posits that pro-drop is the default option in language development; see Section 4.

“early morphosyntactic convergence”), the speakers of non-pro-drop languages seem to set it later (Valian, 1990). That is, they stop omitting subjects at a later stage.

Again, the IA as formalized in Section 2 provides a natural framework for these facts. Non-pro-drop requires the child to acquire the particular grammatical property or rule that blocks pro-drop<sup>9</sup>. What type of evidence leads to positing blocking rules? If the morphosyntactic analysis in Section 2 is on the right track, then impoverishment rules can have this blocking effect. In this case, children would posit them on the basis of evidence from inflectional morphology: there are regularities in the syncretisms across inflectional paradigms which signal rules of impoverishment. We could further conjecture that the assumption that the regularities in verbal paradigms are rule-based and not accidental is reinforced by the observation that as a derivational side-effect, these rules block subject-drop. If children are aware that adult language produces overt subjects where their own grammar (and their discourse-pragmatic knowledge) would allow them to drop subjects (see Section 3.4), positing rules of impoverishment allows them to reach a more target-like production. In other words: impoverishment rules explain two apparently independent properties of adult language. Then, if we were to explain the syntax of English overt subjects on the basis of the overtness condition on Spec, TP that we alluded to in Section 2 (cf. Holmberg, 2010), we would have to appeal for instance to the possibility of *indirect negative evidence* playing a role in acquisition (cf. Chomsky, 1981) and propose that the fact that subjects—and in particular non-referential expressions such as expletives—are systematically overt in adult production supports the assumption that there is a requirement on Spec,TP being overt.

Now, in our analysis, these rules are contingent on the syntax of subjects, and therefore it is to be expected that they will be acquired later than the property making pro-drop possible<sup>10</sup>. Let us take for instance Müller (2006) explanation of the non-pro-drop property in terms of morphological impoverishment. This instance of impoverishment affects the  $\varphi$ -features on T. These features, in turn, result from  $\varphi$ -Agree between T and the subject (cf. Chomsky, 2000, 2001). This means that  $\varphi$ -Agree has to be in place by the time the child learns what the rules of impoverishment of her target language are. Given the implicational relation between Case and Agree (Chomsky, 2000, 2001), we can say that the syntax of subjects, as a whole, precedes the acquisition of the rules of impoverishment. The same dependence with respect to Case and Agree occurs with Holmberg’s (2010) analysis in terms of the overtness requirement on Spec,TP. In order to determine that Spec,TP must be overt, it

<sup>9</sup>Above, following Haegeman and Ihsane (2001), Sato (2011), and Sato and Kim (2012) I suggested that certain varieties of English allow pro-drop. But as discussed by Mack et al. (2012) and Frazier (2015), standard English does not, and the occasional dropping of subjects results from performance factors, where predictable material is reduced. Frazier (2015) highlights that this suggests that the speakers are implicitly aware of the reduction predictable material, and that children may recognize these deviations as being due to the performance system, thus not taking them as evidence that their target is a pro-drop grammar.

<sup>10</sup>The proposal in Jaeggli and Hyams (1988) and Hyams (1991) similarly predicts a later setting of the non-pro-drop option as a result of children realizing late that their target language has poor agreement; see also Section 4.

is necessary to know that it is the subject that is realized there, and that it moves to that position because it Agrees with T. Consequently, with both possible explanations of the non-pro-drop property that we considered in Section 2, it is expected that children will go through a stage in which null subjects are allowed before showing a target-like behavior, where subjects will necessarily be overt. All in all, then, the IA provides a straightforward explanation of what was a rather mysterious consequence of earlier parametric analyses, whereby for instance Italian-speaking children seem to set the parameter relatively earlier than English speakers (see Section 4).

Finally, the impoverishment-based analysis makes a further prediction. Speakers of non-pro-drop languages are expected to take longer than speakers of rich agreement languages before they master verbal inflection. Indeed, acquisition studies show that the production of verbal inflection in early pro-drop languages is virtually errorless and displays higher rates than in early non-pro-drop languages (cf. Hyams, 1991; Phillips, 1996). However, this does not necessarily imply that in the later the inflectional system is not in place: the absence of verbal inflection corresponds in general to the use of root infinitives, and inflected forms, when produced, are also used correctly, which suggests that independent factors could be at play here (cf. Poeppel and Wexler, 1993; Phillips, 1996). More research is thus needed before we can draw conclusions on this issue.

### 3.4. Frequency

The IA characterizes pro-drop as the “default” option. One could think that this directly predicts that the frequency and distribution of null subjects in all early languages should be very similar to that of adult pro-drop languages. However, the IA does not actually make such a prediction. Indeed, pro-drop does not solely depend on structural conditions such as the Case condition discussed above. Completely independent factors also affect the distribution of null vs. overt subjects in the discourse in adult pro-drop languages. For instance, information structure (as mentioned above regarding focus in example 3) and discourse-related factors such as the accessibility or saliency of the antecedent play a crucial role in deciding whether and in what context an argument can be null (Grimshaw and Samek-Lodovici, 1998; Frascarelli, 2007)<sup>11</sup>. Therefore, the process of acquisition of (non-)pro-drop can also only be understood by combining the grammatical level with the discursive-pragmatic level (cf. Hyams and Wexler, 1993 for discussion).

But to what extent do children adhere to discourse conditions on argument omission? Serratrice (2005) shows that like adults, Italian-speaking children tend to realize overtly the arguments that are discursively informative (i.e., those that do not have a salient and accessible antecedent), and to drop those that are uninformative from an early age. Other researchers, such as Clancy (1997) and Allen (2000) obtain comparable results with early Korean and early Inuktitut, respectively.

<sup>11</sup>There are actually many other factors that influence pro-drop that we will not discuss here, such as for instance verb class (cf. Guerriero et al., 2001; Lorusso et al., 2005), or the understanding of the listener’s mental state and perspective (Sorace et al., 2009).

So, both the syntax of Case and the discourse-pragmatic conditions are acquired early. Therefore, the IA predicts that the frequency and distribution of null subjects in all early pro-drop languages should be very similar to that of adult pro-drop languages. And this is indeed confirmed in languages such as Italian (*cf.* Valian, 1990; Lorusso et al., 2005; Serratrice, 2005), Spanish (Bel, 2003), and Catalan (Cabr  Sans and Gavarr , 2006)<sup>12</sup>.

But what about non-pro-drop languages? Does the IA predict that the frequency of null subjects will be the same as in adult pro-drop languages, too? Again, even if pro-drop is syntactically licensed in child languages (due to the early acquisition of the syntax of Case), frequency is also expected to depend on other factors, and in particular on the discourse-pragmatic conditions discussed above. In their study of early English, Hughes and Allen (2006, 2013) report that the more accessible the referent of a subject is, the more likely it is to be null, and the less accessible it is, the more likely it is to be overt, just like in pro-drop languages (see also Guerriero et al., 2001 on later stages of acquisition)<sup>13</sup>.

However, it is well known that the rates of subject-drop in early non-pro-drop languages are much lower than in pro-drop languages. According to Valian (1991), English-speaking children drop subjects at a much lower rate than Italian-speaking children (30% vs. 70%), and Wang et al. (1992) found that the 2-year old English-speaking children in their study showed far fewer null subjects than the Chinese-speaking children (approximately 26% vs. 53%). Under the IA model, null subjects are grammatical in early English. Therefore, the quantitative difference must be explained on independent grounds. What I would like to suggest is the following. English-speaking children, even though they have not yet figured out the grammatical property behind it, are aware of the low frequency (or absence) of null subjects in the adults' grammar. Thus, they produce less null subjects than what the grammar allows (see also Hyams, 1994; O'Grady, 1997 for similar ideas). This is in accordance with the findings in Hughes and Allen (2006, 2013), whereby even though the most highly accessible referents are not always null, they are much more likely to be null than the ones that are less accessible. That is, the discourse-pragmatic factors are comparable to those of Italian, and the patterns are similar, except that overall, the pro-drop option will be appealed to less often.

The difference in the frequency of null subjects between early English and, say, early Chinese or Italian is not something that should surprise us. Variation among adult pro-drop languages is also observed cross-linguistically. For instance Toribio (2000) reports that Dominican Spanish has lower rates of null subjects than Peninsular Spanish, Posio (2012) shows differences between Peninsular Spanish vs. European Portuguese, and Russian can

also be taken to be a pro-drop language that omits subjects at very low rates (McShane, 2005)<sup>14</sup>.

### 3.5. Grammatical Properties of Early Pro-Drop

Besides the timing of acquisition and issues such as the frequency of null subjects, any adequate approach to the early stages of the acquisition of pro-drop should be able to explain the grammatical properties of null subjects in early grammars. Some observations have been made in this regard in the literature, concerning in particular the null subjects produced in early non-pro-drop languages. Some of them are discussed here, arguing that the IA provides a promising framework for their analysis.

#### Expletives

Valian (1991) and Wang et al. (1992) observe that, together with null expletives and null referential subjects, English-speaking children produce overt expletives.

This is expected under the explanation given in Section 3.4 of the higher frequency of overt subjects in early non-pro-drop languages as compared to early pro-drop languages. These children, we have seen, have a pro-drop grammar, which of course allows null expletives. But as a way to converge more closely with the adult's production, where factually, expletives are always overt, they produce less null expletives than what their grammar allows. Note that the alternation between overt and null expletives is not an issue for the claim that early English has a pro-drop grammar, since such patterns are observed in certain adult languages, such as Dominican Spanish (*cf.* Toribio, 2000) and Finnish (*cf.* Holmberg, 2005), which display overt expletives together with null expletives.

#### Root Infinitives

In non-pro-drop languages, null subjects are found mostly in non-finite contexts (*cf.* the overview in Hyams, 2011). How can the IA account for them?

In adult grammars, nonfinite structures can host another type of null subject, standardly referred to as PRO (*cf.* Landau, 2013 for an overview). The first issue is therefore to determine whether the nonfinite null subjects in child grammars are of the *pro*-type or not. Now, in the analysis sketched in Section 2, Case was defined as the condition on pro-drop. Therefore, if we can determine whether in these structures there is a T that assigns Case to its subject, we will be able to characterize the nature of the null subjects they host.

In the early stages of acquisition of non-pro-drop languages, children produce target-deviant constructions with non-finite verbs in root contexts: the so-called root infinitives (or optional infinitives; see Wexler, 2011 for an overview of the literature). Sch tze and Wexler (1996) showed that in English-speaking children's root infinitive structures, about half of the times the (pronominal) subject, if overt, is realized with default accusative case (while in finite contexts the subject is almost always

<sup>12</sup>Some studies report higher frequency of subject omission by children than by adults, which can be explained on independent grounds (*cf.* Serratrice, 2005; Hyams, 2011). For instance, their discourse-situation is often immediate, and their interactions with adults are generally initiated by the latter.

<sup>13</sup>That is, children appear to overgeneralize the use of null subjects when the adult target form would be an overt pronoun or a demonstrative (Hughes and Allen, 2006).

<sup>14</sup>See also Camacho (2013), who proposes that in language change, the first phase of the shift from a pro-drop grammar to a non-pro-drop grammar simply involves an increase in the frequency of overt subject (without there being a change in the syntax).

nominative; see Section 3.2). They take this to indicate absence of Case-assignment to the subject (data from Wexler, 2011, p. 66):

- (5) a. Him fall down. (Nina, 2;3.14, File 17)
- b. Her have a big mouth. (Nina, 2;2.6, File 13)

Root infinitives are among those nonfinite structures where subjects are omitted. Therefore, given that no nominative Case-assignment takes place here, this subject omission does not fall under the analysis put forth here, and will have to be accounted for independently. In fact, it has indeed been proposed that these null subjects are another type of object, possibly PROs (cf. Sano and Hyams, 1994; Bromberg and Wexler, 1995; Schütze and Wexler, 1996; Wexler, 1998)<sup>15</sup>.

### More Finite-Nonfinite Asymmetries

There are some finite contexts in which null subjects are impossible in early non-pro-drop languages. Null subjects are very infrequent with modals (which are inherently finite in English), with finite forms of the copula such as *is*, *am*, *are*, in subordinate clauses or in finite wh-questions (e.g., *Where [e]/he/him going?* vs. *\*Where [e]/he goes?*) (cf. Roeper and Weissenborn, 1990; Valian, 1991; Sano and Hyams, 1994; Bromberg and Wexler, 1995; Roeper and Rohrbacher, 2000). Given the finite nature of the verbs, these cannot be contexts in which the subject is not assigned Case; therefore the explanation will have to be framed in terms of pro-drop being blocked, i.e., there being independent factors that render subject omission impossible. Have children in early stages already learned specifically that agreement on modals and copulas undergoes impoverishment (or that in those constructions SpecTP must be overt, if we adopt Holmberg's, 2010 analysis)? This is highly speculative, but it converges with the observation that even in early pro-drop languages the frequency of subject omission varies with verb class (cf. Guerriero et al., 2001; Allen and Schroeder, 2003; Lorusso et al., 2005). Alternatively, are they postulating another blocking constraint? In this case, what could it be? The non-finiteness restriction on post-wh null subjects, and the impossibility for null subjects in embedded contexts are even more striking: is there something in these CP areas that can block pro-drop?

This is still a poorly understood set of phenomena, and more research is needed before we can make any serious attempt for an explanation. I believe nonetheless that the IA can offer a novel and interesting viewpoint for approaching them. In fact, given that it explains non-pro-drop on the basis of the blocking of pro-drop, it predicts that there may be construction-specific properties in these finite constructions that make subject-drop impossible<sup>16</sup>.

<sup>15</sup>Rizzi (2005a,b) proposes an account that subsumes the null subject phenomenon of early non-pro-drop languages under the root infinitives phenomenon. See Section 4.

<sup>16</sup>An important point that has not been discussed in this paper is that of null objects. The Case condition predicts that objects—to the extent that we assume that they are assigned structural accusative Case—should be omitted during the early stages of acquisition. This is borne out, since object omission occurs both in languages with null objects and in languages without null objects (although at much lower frequencies than subject omission: about 10% in English (Valian, 1991; Wang et al., 1992, 20% in Chinese Wang et al., 1992). A discussion of objects would

## 4. PARAMETER (MIS-)SETTING AND THE INVERSE APPROACH

Hyams (1986) developed a grammar-based approach to the acquisition of (non-)pro-drop which provided support for the Principles and Parameters framework (Chomsky, 1981), arguing that early subject omission in English children's speech was due to the "missetting" of the null subject parameter (more precisely: the *AG/PRO parameter*). The idea is the following. Language acquisition consists in identifying the values of the target language's parameters. Nonetheless, these have a default setting, and the child will change the value of the parameter only if this setting does not account for the input data. In the case of null subjects, Hyams argues, the parameter's default value is positive, which is the value it has in adult languages such as Italian. This explains why early grammars of languages such as English allow pro-drop in a similar way that Italian does.

In following work, Hyams explores the hypothesis that the pro-drop phenomenon is (in part) the by-product of inflectional phenomena, and that null subjects are licensed in early grammar because of the (mis-)setting of a parameterized property of inflection (Jaeggli and Hyams, 1988; Hyams, 1991). More precisely, she adopts Jaeggli and Safir's (1989) analysis of null arguments, whereby null subjects are licensed only in languages with uniformly inflected or uniformly uninflected verbal paradigms, that is, with paradigms composed of complex forms only—i.e., different forms for all person-number combinations, as in Italian—, or with no complex form whatsoever, as in Chinese (the *morphological uniformity principle*)<sup>17</sup>.

Jaeggli and Hyams (1988) and Hyams (1991) propose that null subjects are allowed in early English because children's initial assumption is that the language's morphological paradigm is uniform. Thus, shifting to a non-pro-drop grammar requires them to "realize" that the verbal paradigms are not uniformly inflected.

The analysis of the pro-drop phenomenon proposed in Section 2 shares important aspects with some hypotheses adopted in the parameter (mis-)setting approach; in particular, the idea that the pro-drop phenomenon is (at least in part) the by-product of the properties of inflection. Leaving the theoretical aspects aside (for discussion see Duguine, 2013: chapter 6), what follows discusses the similarities that concern the issue of acquisition. Indeed, in both analyses, early grammars (i) allow pro-drop and (ii) have "uniform" verbal inflectional morphology. Logically then, many predictions made by the IA are also made by Hyams' proposal: null subjects in early pro-drop and non-pro-drop languages, later setting of the non-pro-drop option, dependency of the setting of the

require to first establish assumptions on the nature of a.o. object clitics and clitic optionality in Romance languages, as well as explaining what blocks object-drop in adult languages such as English. I leave these issues open for future research.

<sup>17</sup>With the *morphological uniformity principle*, Jaeggli and Safir (1989) formalize Taraldsen's generalization (cf. Section 2), integrating in the account pro-drop languages that have no verbal agreement morphology (e.g., Japanese and Chinese).



non-pro-drop option on the acquisition of the properties of inflection, etc.<sup>18</sup>.

Different aspects of Hyams' accounts reported above have been challenged conceptually and empirically. What follows discusses three problems that concern basic aspects of Hyams' proposal and shows that adopting the perspective of the IA offers a way to avoid them.

Hyams' parametric approach faces three important issues (see Hyams, 2011 and references therein). First, it does not conform to the Subset Principle, whereby children posit the parameter value that generates the most restricted language consistent with their input data. Indeed, since the positive value of the pro-drop parameter in Italian allows both overt and null subjects, it is a superset of the negative value of English, which only allows overt subjects. Therefore, the Italian setting could not be the initial one. Second, an issue arises with respect to the timing of parameter setting. Data indicates an early setting of the parameter in pro-drop languages, while children acquiring non-pro-drop languages still produce null subjects (see Section 3.3). But if the parameter is set early in Italian, it should also be set early in English. And the third problem is raised by how the explanation based on the morphological uniformity principle is applied to the case of English-like languages. If early English has uniformly uninflected verbal paradigms (just like Chinese), then children beginning to produce inflectional morphology indicates that they have reset their grammar as having non-uniform verbal paradigms. The prediction is thus that they will simultaneously exit the null subject stage. However, this is not borne out: children produce null subjects even after they begin using inflectional morphology.

All three of these issues can be linked to a particular feature of Hyams' analysis: it relies on the existence of a dedicated parameter for pro-drop. As is standardly accepted under the Principles and Parameters framework (*cf.* Chomsky, 1981 and *ff.*), cross-linguistic variation in the availability of pro-drop in the syntax depends on the predetermined set of values of a dedicated parameter. But what if the pro-drop phenomenon was not the (direct) product of a parameter? What if there was no *Null Subject* (or *AGR*, or *morphological uniformity*) *Parameter*?

This is precisely a hypothesis that can be considered and explored under the IA. Indeed, (even) within the Principles and Parameters framework, the model that emerges from the analysis sketched in Section 2 does not conform to that of a parameter, since it postulates that variation in pro-drop emerges from the interplay between different components of the grammar (for discussion see also Duguine et al., in press). It could be considered that the syntax of Case, the rules of impoverishment, and/or the requirement on an overt Spec,TP are parameterized properties. Nonetheless, it has the following features that

distinguish it from standard parametric approaches: (i) pro-drop is universally allowed, and (ii) non-pro-drop is not a core, defining property of languages; it results from pro-drop being systematically blocked in certain configurations in particular languages. This is why, under the IA variation in pro-drop will not be formally characterized as an example of parametric variation.

Crucially, this model will not face the issues that a parametric model such as Hyams' is confronted with. First, the child is complying with the Subset Principle. The IA view does not postulate that acquirers of English posit an incorrect value for a parameter. There is no parameter missetting, and there is no parameter, for that matter. Acquiring a non-pro-drop grammar requires two steps: acquiring the syntax of subjects (*i.e.*, Case) and acquiring the blocking rule. The child makes the first step arguably on the basis of all the morphosyntactic evidence for Case that is available in her primary linguistic data (case morphology, A-movement, etc.). This property is correctly set, that is, it corresponds to the adult grammar. Since pro-drop is universal (to the extent that Case is universal) all children first posit a pro-drop grammar. But even though it is true that English-speaking children's early grammar will generate a language that is a subset of their target language, this is because they have not yet acquired the properties of the grammar that prevent pro-drop. And when doing so, again, they will comply with the Subset Principle, since they will be positing the grammar that generates the most restricted language consistent with their input data.

Second, the IA also allows us to explain the delay in the acquisition of non-pro-drop grammars with respect to pro-drop grammars (see Section 3.3), and predicts that children will attain target-like production progressively, as they acquire the different components of this system, that is, the different linguistic properties that can affect (and in particular, block) pro-drop in the adult language.

Third, if the analysis in terms of impoverishment sketched in Section 2 is on the right track, the production of inflected forms is not expected to correlate with the child exiting the null subject stage. Children will ultimately have to uncover the set of rules of impoverishment affecting inflectional morphemes before they stop dropping subjects, that is, they will have to realize that the syncretisms in verbal paradigms are not accidental, that they result from morphological rules (which, incidentally, block pro-drop; see Section 3.3). That is, conceivably, until that point they can produce inflected or non-inflected forms and null subjects.

To conclude, the approach explored in this paper offers a perspective on the acquisition of pro-drop that shares important features with earlier work, in particular Hyams (1986, 1991), and makes various similar predictions. However, in contrast with these, it also implies that there is no Pro-Drop Parameter as such. The patterns of null/overt subjects across languages emerge from the conspiracy between different components of the grammar, and the stages of language development result from the timing in the acquisition of these components. This difference allows it to circumvent some problems that Hyams' proposal was confronted to. So, the IA can be

<sup>18</sup>Note that even though in Hyams (1986), Jaeggli and Hyams (1988), and Hyams (1991) pro-drop is the default option in language development, it is not the default option in the syntax, in the sense considered in Section 2. Indeed, these analyses assume that the syntax of pro-drop is constrained by specific grammatical mechanisms (licensing and identification conditions), which implies that conceptually non-pro-drop is the default option.

seen as the opportunity to re-open the discussion on the acquisition of null arguments, and explore with new tools an account that was quite well supported both conceptually and empirically.

It must be pointed out that in the years following Hyams' work, studies showed that there are differences in the distribution of null subjects in early non-pro-drop languages vs. early pro-drop languages. These concern observations that were cited in Section 3.5: null subjects are very infrequent with modals, with finite forms of the copula such as *is*, *am*, *are*, in subordinate clauses or in finite wh-questions (e.g., *Where [e]/he/him going?* vs. *\*Where [e]/he goes?*) (cf. Roeper and Weissenborn, 1990; Valian, 1991; Sano and Hyams, 1994; Bromberg and Wexler, 1995; Roeper and Rohrbacher, 2000). This observation, combined with other issues such as the three points discussed above, have led many researchers to consider that early missing subjects in non-pro-drop languages are not part of the pro-drop phenomena. In particular, Rizzi (2005a,b) develops an influential account whereby null subjects in early non-pro-drop languages result from "root subject drop," a (parameterized) grammatical option where the specifier of root/truncated clauses (bare IPs) can be null. The root subject drop analysis straightforwardly accounts for "root" effects in early English such as the impossibility for null subjects to occur after a wh-phrase or in a subordinate clause, but as noted by Hyams (2011), it does not explain why they do not occur with modals (Valian, 1991) or with finite forms of the copula (Sano and Hyams, 1994)<sup>19</sup>. Section 3.5 merely sketched a possible explanation for the latter facts under the IA, but I hope to have shown that, even though much remains to be done, the IA can be seen as a version of the parameter missetting approach, which circumvents some earlier problems, and which can be investigated as an alternative explanation of the acquisition of (so-called) pro-drop vs. non-pro-drop languages.

<sup>19</sup>Hyams (2011, pp. 27–30) also raises more general questions (see also Serratrice and Allen, 2015). For instance, whether Italian children also posit the positive value of the parameter that makes root subject drop possible, and how the two parameters are expected to interact.

## REFERENCES

- Alexiadou, A., and Anagnostopoulou, E. (1998). Parametrizing Agr: word order, V-Movement and EPP-Checking. *Nat. Lang. Linguist. Theory* 16, 491–539. doi: 10.1023/A:1006090432389
- Allen, S. E. (2000). A discourse-pragmatic explanation for argument representation in child Inuktitut. *Linguistics* 38, 483–521. doi: 10.1515/ling.38.3.483
- Allen, S. E., and Schroeder, H. (2003). "Preferred argument structure in early Inuktitut spontaneous speech data," in *Preferred Argument Structure: Grammar as Architecture for Function*, eds J. W. D. Bois, L. E. Kumpf, and W. J. Ashby (Amsterdam: John Benjamins), 301–338.
- Barbosa, P. (1995). *Null Subjects*. Ph.D. thesis, MIT.
- Barreña, A. (1995). *Haur Euskaldunen Hizkuntzaren Jabekuntza-Garapena: INFL eta KONP Funtzio-Kategorien Erabileraz*. Bilbao: University of the Basque Country.
- Bel, A. (2003). The syntax of subjects in the acquisition of Spanish and Catalan. *Probus Int. J. Latin Romance Linguist.* 15, 1–26. doi: 10.1515/prbs.2003.003
- Bennis, H. (2006). "Agreement, *pro* and imperatives," in *Arguments and Agreement*, eds P. Ackema, P. Brandt, M. Schoorlemmer, and F. Weermann (Oxford: Oxford University Press), 101–123.
- Bonet, E. (1991). *Morphology after Syntax*. Ph.D. thesis, MIT.
- Bromberg, H., and Wexler, K. (1995). "Null subjects in wh-questions," in *MIT Working Papers Linguistics*, Vol. 26 (Cambridge, MA), 221–248.
- Cabré Sans, Y., and Gavarró, A. (2006). "Subject distribution and verb classes in child Catalan," in *Proceedings of the 2nd Conference of GALANA* (Somerville, MA: Cascadilla Press), 51–60.
- Cai, Z. G., Pickering, M. J., Wang, R., and Branigan, H. P. (2014). It is there whether you hear it or not: syntactic representation of missing arguments. *Cognition* 136, 255–267. doi: 10.1016/j.cognition.2014.11.017
- Camacho, J. (2013). *Null Subjects*. Cambridge: Cambridge University Press.
- Cardinaletti, A., and Starke, M. (1999). "The typology of structural deficiency: a case study of three classes of pronouns," in *Clitics in the Languages of Europe*, ed H. Van Riemsdijk (Berlin: Mouton de Gruyter), 145–253.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Cambridge: Mouton de Gruyter.

## 5. CONCLUSIONS

The *inverse approach to pro-drop* (IA) proposes a shift in the way the question of pro-drop is addressed. This paper has focused on its implications for language development, showing that it offers an explanatory account of several properties of child languages, both with pro-drop and non-pro-drop target languages. It shares important features with earlier proposals, such as in particular the parameter missetting approach developed in Hyams (1986, 1991). But there are also crucial differences. In particular, the conceptual consequence that there is no Pro-Drop Parameter as such allows us to circumvent some issues raised by these earlier accounts. In other words, the results obtained here suggest the parameter missetting approach can be brought back to the research in the acquisition of (non-)pro-drop, since they give a way to formalize the developmental intuition that all children start out with a pro-drop grammar, and that this is why those speaking a pro-drop language will show early target-like behavior and those speaking a non-pro-drop language will shift to a different grammar later on.

Much remains to be done, but I hope the above discussion succeeds in showing that the IA opens some avenues of research that are worth exploring, and which might give new solutions to old problems.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## FUNDING

The present work was made possible through funding from the Basque Government (IT769-13), the Spanish Ministry of Economy and Competitiveness (FFI2014-53675-P, FFI2014-51675REDT), and the University of the Basque Country UPV/EHU (UFI11/14).

- Chomsky, N. (1982). *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, MA: MIT Press.
- Chomsky, N. (2000). "Minimalist inquiries: the framework," in *Step by Step: Paper in Honor of Howard Lasnik*, eds R. Martin, D. Michaels, and J. Uriagereka (Cambridge, MA: MIT Press), 89–155.
- Chomsky, N. (2001). "Derivation by phase," in *Ken Hale: A Life in Language*, ed M. Kenstowicz (Cambridge, MA: MIT Press), 1–52.
- Clancy, P. (1997). "Discourse motivations of referential choice in Korean acquisition," in *Japanese/Korean linguistics*, Vol. 6, eds H. Sohn and J. Haig (Stanford, CA: CSLI Publications), 639–659.
- Dryer, M. S. (2013). "Expression of pronominal subjects," in *The World Atlas of Language Structures Online*, eds M. S. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology). Available online at: <http://wals.info/chapter/101> (Accessed January 19, 2017).
- Duguine, M. (2013). *Null Arguments and Linguistic Variation: A Minimalist Analysis of Pro-drop*. Ph.D. thesis, Université de Nantes and UPV/EHU.
- Duguine, M. (2014). Argument ellipsis: a unitary approach to pro-drop. *Linguist. Rev.* 31, 515–550. doi: 10.1515/ldr-2014-0010
- Duguine, M., Irurtzun, A., and Boeckx, C. (in press). Linguistic diversity and granularity: two case-studies against parametric approaches. *Linguist. Anal.*
- Ezeizabarrena, M. J. (2002). "Root infinitives in two pro-drop languages," in *The Acquisition of Spanish Morphosyntax*, eds A. T. Pérez-Leroux and J. M. Licerias (Dordrecht: Kluwer), 35–65.
- Frampton, J. (2002). "Syncretism, impoverishment, and the structure of person features," in *Papers from the Chicago Linguistics Society Meeting*, Vol. 38, eds M. Andronis, E. Debenport, A. Pycha, and K. Yoshimura (Chicago), 207–222.
- Frascarelli, M. (2007). Subjects, topics and the interpretation of referential pro. *Nat. Lang. Linguist. Theory* 25, 691–734. doi: 10.1007/s11049-007-9025-x
- Frazier, L. (2015). Do null subjects (mis-)trigger pro-drop grammars? *J. Psycholinguist. Res.* 44, 669–674. doi: 10.1007/s10936-014-9312-8
- Grimshaw, J., and Samek-Lodovici, V. (1998). "Optimal subjects and subject universals," in *Is the Best Good Enough? Optimality and Competition in Syntax*, eds P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, and D. Pesetsky (Cambridge, MA: MIT Press), 193–219.
- Guasti, M.-T. (1993/1994). Verb syntax in Italian child grammar: finite and non-finite verbs. *Lang. Acquisit.* 3, 1–40. doi: 10.1207/s15327817la0301\_1
- Guerrierio, A. S., Cooper, A., Oshima-Takane, Y., and Kuriyama, Y. (2001). "A discourse-pragmatic explanation for argument realization and omission in English and Japanese children's speech," in *Proceedings of the 25th annual Boston University Conference on Language Development*, eds A. H.-J. Do, L. Dominguez, and A. Johansen (Somerville, MA: Cascadilla Press), 319–330.
- Haegeman, L., and Ihsane, T. (2001). Adult null subjects in the non-pro drop languages: two diary dialects. *Lang. Acquisit.* 9, 329–346. doi: 10.1207/S15327817LA0904\_03
- Halle, M. (1997). "Distributed morphology: impoverishment and fission," in *MIT Working Papers Linguistics*, Vol. 30 (Cambridge, MA), 425–449.
- Harley, H., and Noyer, R. (1999). State-of-the-article: distributed morphology. *GLOT Int.* 4, 3–9.
- Hoekstra, T., and Hyams, N. (1998). Aspects of root infinitives. *Lingua* 106, 81–112. doi: 10.1016/S0024-3841(98)00030-8
- Holmberg, A. (2005). Is there a little pro? Evidence from Finnish. *Linguist. Inq.* 36, 533–564. doi: 10.1162/002438905774464322
- Holmberg, A. (2010). "Null subject parameters," in *Parametric Variation*, eds T. Biberauer, A. Holmberg, I. Roberts, and M. Sheehan (Cambridge: Cambridge University Press), 88–124.
- Huang, C. (1984). On the distribution and reference of empty pronouns. *Linguist. Inq.* 15, 531–574.
- Hughes, M. E., and Allen, S. E. (2006). "A discourse-pragmatic analysis of subject omission in child English," in *Proceedings of the 30th Annual Boston University Conference on Language Development*, eds D. Bamman, T. Magnitskaia, and C. Zaller (Somerville, MA: Cascadilla Press), 293–304.
- Hughes, M. E., and Allen, S. E. (2013). The effect of individual discourse-pragmatic features on referential choice in child English. *J. Pragmat.* 56, 15–30. doi: 10.1016/j.pragma.2013.05.005
- Hyams, N. (1986). *Language Acquisition and the Theory of Parameters*. Dordrecht: Foris.
- Hyams, N. (1991). "A reanalysis of null subjects in child language," in *Theoretical Issues in Language Acquisition*, eds H. G. Jürgen Weissenborn and T. Roeper (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.)
- Hyams, N. (1994). "Commentary: null subjects in child language and the implications of cross-linguistic variation," in *Syntactic Theory and First Language Acquisition: Cross-Linguistic Perspectives, Vol. 2: Binding, Dependencies, and Learnability*, eds B. Lust, G. Hermon, and J. Kornfilt (Hillsdale, NJ: Lawrence Erlbaum), 287–299.
- Hyams, N. (2011). "Missing subjects in early child language," in *Handbook of Generative Approaches to Language Acquisition*, eds J. de Villiers and T. Roeper (Dordrecht: Springer), 13–52.
- Hyams, N., Mateu, V., Orfitelli, R., Putnam, M., Rothman, J., and Sánchez, L. (2015). "Parameters in language acquisition and contact," in *Contemporary Handbook of Linguistic Parameters*, eds A. Fábregas, J. Mateu, and M. Putnam (London: Bloomsbury), 353–375.
- Hyams, N., and Wexler, K. (1993). On the grammatical basis of null subjects in child language. *Linguist. Inq.* 24, 421–459.
- Jaeggli, O., and Hyams, N. (1988). "Morphological uniformity and the setting of the null subject parameter," in *Proceedings of the NELS 18* (Amherst, MA: University of Massachusetts, GLSA), 239–253.
- Jaeggli, O., and Safir, K. J. (eds.). (1989). "The null subject parameter and parametric theory," in *The Null Subject Parameter* (Dordrecht: Kluwer), 185–214.
- Kim, Y.-J. (1997). "The acquisition of Korean," in *The Cross-linguistic Study of Language Acquisition*, Vol. 4, ed D. I. Slobin (Hillsdale, NJ: Lawrence Erlbaum), 335–435.
- Landau, I. (2013). *Control in Generative Grammar: A Research Companion*. Cambridge: Cambridge University Press.
- Larson, R. K., and Luján, M. (1989). *Emphatic Pronouns*. New York, NY; Austin, TX: Stony Brook University; University of Texas.
- Legate, J. (2006). "Split absolutive," in *Ergativity*, ed A. Johns (Amsterdam: Springer), 143–171.
- Lorusso, P., Caprin, C., and Guasti, M.-T. (2005). "Overt subject distribution in early Italian children," in *A Supplement to the Proceedings of the Twenty-ninth Boston University Conference on Language Development*, eds A. Brugos, M. R. Clark-Cotton, and S. Ha (Somerville, MA: Cascadilla Press).
- Mack, J., Clifton, C. J., Frazier, L., and Taylor, P. V. (2012). (Not) hearing optional subjects: the effects of pragmatic usage preferences. *J. Mem. Lang.* 67, 211–223. doi: 10.1016/j.jml.2012.02.011
- McCloskey, J., and Hale, K. (1984). On the syntax of person-number inflection in Modern Irish. *Nat. Lang. Linguist. Theory* 1, 487–533.
- McShane, M. (2005). *Asubjects Ellipsis*. Oxford: Oxford University Press.
- Müller, G. (2006). "Pro-drop and impoverishment," in *Form, Structure, and Grammar. A Festschrift Presented to Gunther Grewendorf on Occasion of His 60th Birthday*, eds P. Brandt and E. Fuß (Berlin: Akademie Verlag), 93–115.
- Müller, G. (2008). "Some consequences of an impoverishment-based approach to morphological richness and pro-drop," in *Elements of Slavic and Germanic Grammars: A Comparative View*, eds J. Witkos and G. Fanselow (Frankfurt: Lang), 125–145.
- O'Grady, W. (1997). *Syntactic Development*. Chicago, IL: University of Chicago Press.
- Phillips, C. (1996). "Root infinitives are finite," in *Proceedings of the 20th Annual Boston University Conference on Language Development*, eds A. Stringfellow, D. Cahana-Amitay, E. Hughes, and A. Zukowski (Somerville, MA: Cascadilla Press), 588–599.
- Pierce, A. (1992). *Language Acquisition and Syntactic Theory - A Comparative Analysis of French and English Child Grammars*. Dordrecht: Kluwer.
- Platzack, C. (1994). "Null subjects, weak Agr and syntactic differences in Scandinavian," in *Working Papers in Scandinavian Syntax*, Vol. 43 (Lund), 51–76.
- Platzack, C., and Holmberg, A. (1989). "The role of AGR and finiteness," in *Working Papers in Scandinavian Syntax*, Vol. 44 (Lund), 101–117.
- Poeppl, D., and Wexler, K. (1993). The full competence hypothesis of clause structure in early German. *Language* 69, 1–33. doi: 10.2307/416414
- Posio, P. (2012). *Pronominal Subjects in Peninsular Spanish and European Portuguese. Semantics, Pragmatics, and Formulaic Sequences*. Ph.D. thesis, University of Helsinki.

- Raposo, E. (1986). "The null object in European Portuguese," in *Studies in Romance Linguistics*, ed O. Jaeggli and C. Silva-Corvalán (Dordrecht: Foris), 373–390.
- Rizzi, L. (1982). *Issues in Italian Syntax*. Dordrecht: Foris.
- Rizzi, L. (1986). Null objects in Italian and the theory of *pro*. *Linguist. Inq.* 17, 501–557.
- Rizzi, L. (2005a). "Grammatically-based target-inconsistencies in child language," in *Proceedings of the Inaugural Conference of GALANA, UCONN/MIT Working papers in Linguistics*, eds K. U. Deen, J. Nomura, B. Schulz, and B.D. Schwartz (Cambridge, MA), 19–49.
- Rizzi, L. (2005b). "On the grammatical basis of language development: a case study," in *The Oxford handbook of comparative syntax*, eds G. Cinque and R. Kayne (Oxford: Oxford University Press), 70–109.
- Roberge, Y. (1990). *Syntactic Recoverability of Null Arguments*. Montréal, QC: McGill-Queen's Press - MQUP.
- Roberts, I. (1993). *Verbs and Diachronic Syntax: A Comparative History of English and French*. Dordrecht: Kluwer.
- Roberts, I. (2010a). "A deletion analysis of null subjects," in *Parametric variation*, eds T. Biberauer, A. Holmberg, I. Roberts, and M. Sheehan (Cambridge: Cambridge University Press), 58–87.
- Roberts, I. (2010b). "Varieties of French and the null subject parameter," in *Parametric Variation*, eds T. Biberauer, A. Holmberg, I. Roberts, and M. Sheehan (Cambridge: Cambridge University Press), 303–327.
- Roeper, T., and Rohrbacher, B. (2000). "True Pro-drop in child English and the principle of economy or projection," in *The Acquisition of Scrambling and Cliticization*, eds S. Powers and C. Hamann (Dordrecht: Kluwer), 345–396.
- Roeper, T., and Weissenborn, J. (1990). "How to make parameters work: comments on Valian," in *Language Processing and Language Acquisition*, eds L. Frazier and J. De Villiers (Dordrecht: Kluwer), 147–162.
- Rohrbacher, B. (1999). *Morphology-Driven Syntax: A Theory of V-to-I Raising and Pro-drop*. Amsterdam: J. Benjamins.
- Rosenkvist, H. (2009). "Null referential subjects in Germanic languages an overview," in *Working Papers in Scandinavian Syntax*, Vol. 84 (Lund), 151–180.
- Saito, M. (2007). Notes on East Asian argument ellipsis. *Lang. Res.* 43, 203–227.
- Sano, T., and Hyams, N. (1994). "Agreement, finiteness, and the development of null arguments," in *Proceedings of NELS 24*, ed M. Gonzáles (Amherst, MA: GLSA), 543–548.
- Sato, Y. (2011). Radical pro-drop and fusional pronominal morphology in Colloquial Singapore English: reply to Neeleman and Szendrői. *Linguist. Inq.* 42, 356–365. doi: 10.1162/LING\_a\_00046
- Sato, Y., and Kim, C. (2012). Radical pro-drop and the role of syntactic agreement in Colloquial Singapore English. *Lingua* 122, 858–873. doi: 10.1016/j.lingua.2012.02.006
- Schütze, C. (2001). On the nature of default case. *Syntax* 4, 205–238. doi: 10.1111/1467-9612.00044
- Schütze, C., and Wexler, K. (1996). "Subject case licensing and English Root Infinitives," in *Proceedings of the 20th Annual Boston University Conference on Language Development* (Somerville, MA: Cascadilla Press), 670–681.
- Serratrice, L. (2005). The role of discourse pragmatics in the acquisition of subjects in Italian. *Appl. Psycholinguist.* 26, 437–462. doi: 10.1017/S0142716405050241
- Serratrice, L., and Allen, S. E. (2015). "The acquisition of reference: introduction," in *The Acquisition of Reference*, eds L. Serratrice and S. E. Allen (Amsterdam: John Benjamins), 1–24.
- Sorace, A., Serratrice, L., Filici, F., and Baldo, M. (2009). Discourse conditions on subject pronoun realization: testing the linguistic intuitions of older bilingual children. *Lingua* 119, 460–477. doi: 10.1016/j.lingua.2008.09.008
- Speas, M. (2006). "Economy, agreement, and the representation of null arguments," in *Arguments and Agreement*, eds P. Ackema and F. Weerman (Oxford: Oxford University Press), 35–75.
- Taraldsen, T. (1980). *On the Nominative Island Condition, Vacuous Application and the That-Trace Filter*. Bloomington, IN: Indiana University Linguistics Club.
- Toribio, A. J. (2000). Setting parametric limits on dialectal variation in Spanish. *Lingua* 10, 315–341. doi: 10.1016/S0024-3841(99)00044-3
- Vainikka, A., and Levy, Y. (1999). Empty subjects in Finnish and Hebrew. *Nat. Lang. Linguist. Theory* 17, 613–671. doi: 10.1023/A:1006225032592
- Valian, V. (1990). A problem for parameter-setting models of language acquisition. *Cognition* 35, 105–122. doi: 10.1016/0010-0277(90)90011-8
- Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition* 40, 21–81. doi: 10.1016/0010-0277(91)90046-7
- Valian, V., and Eisenberg, Z. (1996). The development of syntactic subjects in Portuguese-speaking children. *J. Child Lang.* 23, 103–128. doi: 10.1017/S0305000900010114
- Wang, Q., Lillo-Martin, D., Best, C. T., and Levitt, A. (1992). Null subject versus null object: some evidence from the acquisition of Chinese and English. *Lang. Acquisit.* 2, 221–254. doi: 10.1207/s15327817la0203\_2
- Wexler, K. (1998). Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. *Lingua* 106, 23–79. doi: 10.1016/S0024-3841(98)00029-1
- Wexler, K. (2011). "Grammatical computation in the Optional Infinitive stage," in *Handbook of Generative Approaches to Language Acquisition*, eds J. de Villiers and T. Roeper (Dordrecht: Springer), 53–118.
- Zribi-Hertz, A. (1994). "La syntaxe des clitics nominatifs en français standard et en français avancé," in *Travaux de Linguistique et Philologie*, eds G. Kleiber and G. Roques (Strasbourg: Klincksieck), 131–147.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer EP and the handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Duguine. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Dissociating Effects of Scrambling and Topicalization within the Left Frontal and Temporal Language Areas: An fMRI Study in Kaqchikel Maya

OPEN ACCESS

Shinri Ohta<sup>1,2†</sup>, Masatoshi Koizumi<sup>3</sup> and Kuniyoshi L. Sakai<sup>1,2\*</sup>

**Edited by:**

Ángel J. Gallego,  
Autonomous University of Barcelona,  
Spain

**Reviewed by:**

John E. Drury,  
Stony Brook University, USA  
Kepa Erdocia,  
University of the Basque Country,  
Spain

**\*Correspondence:**

Kuniyoshi L. Sakai  
sakai@mind.c.u-tokyo.ac.jp

**† Present address:**

Shinri Ohta,  
Department of Linguistics, Faculty  
of Humanities, Kyushu University,  
6-19-1 Hakozaki, Higashi-ku,  
Fukuoka-shi, Fukuoka, Japan

**Specialty section:**

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 19 November 2016

**Accepted:** 25 April 2017

**Published:** 09 May 2017

**Citation:**

Ohta S, Koizumi M and Sakai KL  
(2017) Dissociating Effects  
of Scrambling and Topicalization  
within the Left Frontal and Temporal  
Language Areas: An fMRI Study  
in Kaqchikel Maya.  
*Front. Psychol.* 8:748.  
doi: 10.3389/fpsyg.2017.00748

<sup>1</sup> Department of Basic Science, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan, <sup>2</sup> Core Research for Evolutionary Science and Technology, Japan Science and Technology Agency, Tokyo, Japan, <sup>3</sup> Department of Linguistics, Graduate School of Arts and Letters, Tohoku University, Sendai, Japan

Some natural languages grammatically allow different types of changing word orders, such as object scrambling and topicalization. Scrambling and topicalization are more related to syntax and semantics/phonology, respectively. Here we hypothesized that scrambling should activate the left frontal regions, while topicalization would affect the bilateral temporal regions. To examine such distinct effects in our functional magnetic resonance imaging study, we targeted the Kaqchikel Maya language, a Mayan language spoken in Guatemala. In Kaqchikel, the syntactically canonical word order is verb-object-subject (VOS), but at least three non-canonical word orders (i.e., SVO, VSO, and OVS) are also grammatically allowed. We used a sentence-picture matching task, in which the participants listened to a short Kaqchikel sentence and judged whether a picture matched the meaning of the sentence. The advantage of applying this experimental paradigm to an understudied language such as Kaqchikel is that it will allow us to validate the universality of linguistic computation in the brain. We found that the conditions with scrambled sentences [+scrambling] elicited significant activation in the left inferior frontal gyrus and lateral premotor cortex, both of which have been proposed as grammar centers, indicating the effects of syntactic loads. In contrast, the conditions without topicalization [–topicalization] resulted in significant activation in bilateral Heschl's gyrus and superior temporal gyrus, demonstrating that the syntactic and phonological processes were clearly dissociated within the language areas. Moreover, the pre-supplementary motor area and left superior/middle temporal gyri were activated under relatively demanding conditions, suggesting their supportive roles in syntactic or semantic processing. To exclude any semantic/phonological effects of the object-subject word orders, we performed direct comparisons while making the factor of topicalization constant, and observed localized activations in the left inferior frontal gyrus and lateral premotor cortex. These results establish that the types of scrambling

and topicalization have different impacts on the specified language areas. These findings further indicate that the functional roles of these left frontal and temporal regions involve *linguistic* aspects themselves, namely syntax versus semantics/phonology, rather than output/input aspects of speech processing.

**Keywords:** language, syntax, word order, scrambling, topicalization, inferior frontal gyrus, lateral premotor cortex, fMRI

## INTRODUCTION

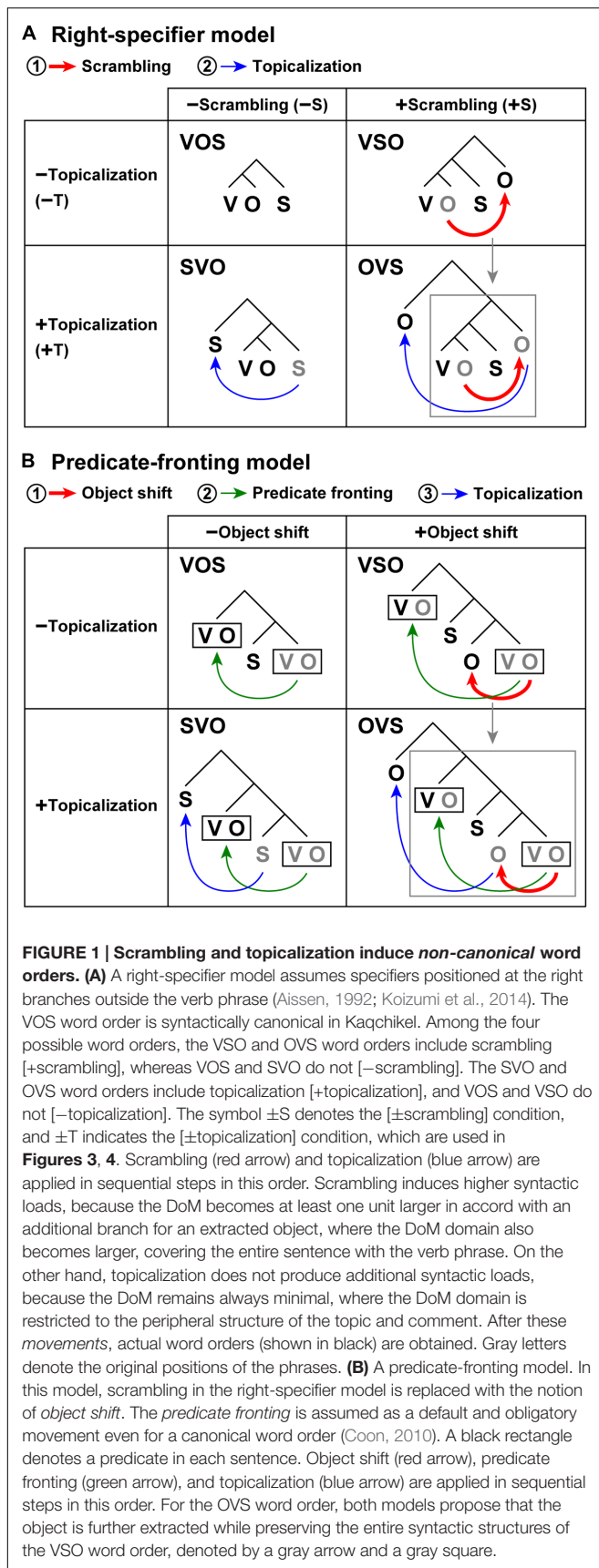
There are natural languages that grammatically allow different types of changing word orders (Karimi, 2003). This phenomenon can be explained by *movement* of phrases, which is a key operation proposed in modern linguistics. A word order with the simplest syntactic structure is *syntactically canonical*, and word orders that are a result of a movement of phrases are non-canonical. The notion of such canonicity, as well as syntactic knowledge, is independent of the frequency/probability of usage, or learning of words (Chomsky, 1957). One type of movement is *object scrambling*, where an object (O) to be emphasized is extracted from the original position in a verb phrase and moved to a structurally higher position, skipping other phrases and resulting in more complex tree structures. In this article, we refer to “object scrambling” as simply “scrambling.” Scrambling is not allowed in English, but scrambled sentences are grammatical in Japanese. Although there are information structure distinctions (e.g., emphasis) related to scrambling, scrambling in Japanese does not change the grammatical relations (e.g., subject, direct object, and indirect object) and semantic roles (e.g., agent, patient, and experiencer) of a sentence (Fukui, 1993; Saito and Fukui, 1998).

Our previous study using functional magnetic resonance imaging (fMRI) revealed selective responses to scrambled sentences in the left frontal regions: the opercular and triangular parts (L. F3op/F3t) of the left inferior frontal gyrus (IFG), as well as the left lateral premotor cortex (L. LPMC) (Kinno et al., 2008). In our magnetoencephalography (MEG) study (Inubushi et al., 2012), we observed the effects of canonicity in the left IFG in response to more complex ditransitive sentences (i.e., those including a verb and two objects). We also demonstrated that the Degree of Merger (DoM) accounted for syntax-selective activations in the L. F3op/F3t (Ohta et al., 2013b). The DoM is the maximum depth of merged subtrees (i.e., Mergers) within an entire sentence, and it properly measures the complexity of tree structures. The DoM domain, i.e., the subtrees where the DoM is calculated, is an entire sentence when there is no constraint, but this changes dynamically in accord with syntactic operations and/or task requirements (Ohta et al., 2013a). Scrambling induces higher syntactic loads, because the DoM becomes at least one unit larger in accord with an additional branch for an extracted object, where the DoM domain also becomes larger covering entire sentences with a verb phrase (see Figure 6 of Ohta et al., 2013a). In addition to the L. F3op/F3t and L. LPMC, some fMRI and MEG studies have proposed that the left anterior temporal lobe (L. ATL) is also

specialized in the construction of complex meaning (Poeppel et al., 2012), although effects of a movement of phrases have not been previously examined by those studies. By directly contrasting scrambled sentences with non-scrambled ones, the relative contribution of the L. F3op/F3t and L. ATL should be clarified.

Another type of movement is *topicalization*, in which, for example, a subject (S) or an object outside a verb phrase moves to a still structurally higher position to represent a topic, i.e., information that has already been mentioned in the discourse/context. In English, topicalization of an object generates a *non-canonical* word order (Radford, 2009), such as “*John read a book. That book, Mary read at school,*” in which a two-step movement of an object is involved. Here, the given information is presented as a topic at the initial position of the sentence, which makes sentence comprehension easier. Indeed, a sentence with the same movement of a non-topic noun phrase becomes ungrammatical: “\**A book, Mary read at school.*” In the absence of topicalization, semantic/phonological loads and general auditory attention would become larger, because *all* words should be attended without prior information, rather than a particular topicalized word. Another possibility is that a topicalized sentence becomes semantically and phonologically marked, which may increase semantic/phonological loads in comparison with the canonical word order. By examining both effects of [ $\pm$ topicalization] in brain activation, we would be able to determine which of these effects is more prominent.

While topicalization and scrambling are inseparable in rigid word-order languages such as English and Hebrew, they become separable in flexible word-order languages like Japanese. In the latter case, the DoM domain can be restricted to the peripheral structure of the topic and comment (i.e., the rest of the sentence), and thus topicalization does not produce additional syntactic loads, because the DoM remains minimal. An ERP study using topicalization and *wh*-questions in German reported that both constructions elicited a left-anterior negativity, which is typically interpreted as indexing an increase in memory burden (Felser et al., 2003). However, in German topicalized sentences, any effects due to a two- or multiple-step movement of an object should be considered, where the first-step of such a movement involves scrambling just as OVS in Kaqchikel (see Figure 1A). Moreover, topicalization may have enhanced memory burden, since no specific context was provided for each presented sentence in that study. Scrambling and topicalization are thus more related to syntax and semantics/phonology, respectively. According to psycholinguistic studies, the differences between these two types of movements do not seem to affect behavioral



data (Sekerina, 2003). The use of fMRI would dissociate the effects of these movements among multiple language areas. Our previous studies have clarified that syntactic processing, i.e., movement or merger of phrases, activates the L. F3op/F3t (Kinno et al., 2008; Ohta et al., 2013a,b), while phonological loads and auditory attention activates the bilateral superior temporal gyrus (STG) (Suzuki and Sakai, 2003). Based on these studies, we hypothesized that the main effects of scrambling should activate the L. F3op/F3t and L. LPMC, while the main effects of topicalization would affect the bilateral temporal regions.

To dissociate the effects of [+scrambling] and [±topicalization], a flexible word-order language that grammatically allows four different word orders, i.e., [±scrambling, ±topicalization], should be targeted, which can be realized with Kaqchikel Maya (hereafter “Kaqchikel”), a Mayan language spoken in Guatemala. To our knowledge, the present study is the first to examine that [±scrambling] and [±topicalization] can be separated symmetrically within participants, sessions, and a language. In Kaqchikel, the syntactically canonical word order is verb-object-subject (VOS), but at least three non-canonical word orders (i.e., SVO, VSO, and OVS) are also grammatically allowed (**Figure 1**; García Matzar and Rodríguez Guaján, 1997; Brown et al., 2006). Previous neuroimaging and psycholinguistic studies have mainly targeted *SO languages*, where the S precedes the O in a canonical word order (e.g., SVO and SOV), such as English, Japanese, and German. Sentences with the non-canonical OS word order (e.g., scrambling) are more difficult to process than those with the canonical word order, while keeping all other factors such as semantic roles equal (Marantz, 2005). Indeed, fMRI studies have reported increased activation by non-canonical word orders in the left IFG (Bahlmann et al., 2007; Kinno et al., 2008; Kim et al., 2009), which may reflect the effects of scrambling. A neuroimaging study has described the enhanced neural effects of topicalization (Ben-Shachar et al., 2004), in which a two-step movement of an object is involved in Hebrew sentences as in English and German. Because no specific context was provided for each presented sentence in that latter study, it is also possible that topicalization artificially enhanced syntactic and semantic/phonological processes. Such an activation increase might be triggered by the OS word order itself, which is related to one of “irregular prominence factors of noun phrases,” such as Patient vs. Agent, Inanimate vs. Animate, etc. (Bornkessel et al., 2005; Grewe et al., 2006). To conclusively examine which of these accounts is correct in fMRI experiments, we have targeted the *OS language* of Kaqchikel, where the O precedes the S in a *canonical* word order. If the last possibility is correct, and the activation increase is triggered by the OS word order itself, then the canonical word order in Kaqchikel (VOS) would elicit higher activations than the non-canonical word order (VSO), which seems unlikely. We predict that VSO elicits higher activations than VOS.

Kaqchikel is a *head-marking language*, in which prefixes of a *verb* (i.e., a head in a sentence) specify numbers (singular or plural) and persons (first, second, or third) of the object/subject,

whereas English, Japanese, and German are *dependent-marking languages*, in which noun phrases (dependents) that depend on a verb are always marked for subjects and objects when possible (like English pronouns). Regarding Mayan sentences, a right-specifier model has been proposed for the syntactic structures of a sentence (Aissen, 1992; Koizumi et al., 2014), assuming specifiers positioned at the right branches outside the verb phrase VO, in addition to specifiers of a complementizer phrase (e.g., “that”) positioned at the left branches (Figure 1A). For the canonical word order (VOS), the S is a specifier positioned at a right branch. Moreover, for the VSO and OVS word orders, scrambling of the O results in a right-specifier. On the other hand, for the SVO and OVS word orders, topicalization of the S or O results in a left-specifier of a complementizer phrase. These four word orders thus have the following factors: VOS [–scrambling, –topicalization], SVO [–scrambling, +topicalization], VSO [+scrambling, –topicalization], OVS [+scrambling, +topicalization]. Another linguistic study has proposed an alternative model, i.e., a predicate-fronting model (Figure 1B; Coon, 2010), which is basically consistent with the right-specifier model and will be discussed later.

Based on our earlier investigations (Kinno et al., 2008, 2014), we used here a modified sentence-picture matching task, in which each participant listened to a Kaqchikel sentence and judged whether a picture matched the meaning of the sentence (Figure 2). The advantage of applying this experimental paradigm to an understudied language such as Kaqchikel is that it will allow us to validate the universality of linguistic computation in the brain.

## BASICS OF KAQCHIKEL SYNTAX

Kaqchikel is an *ergative language*, in which a subject of a transitive verb is marked by an *ergative case*, whereas an object of a transitive verb, as well as a subject of an intransitive verb, is marked by an *absolutive case*; here we used transitive verbs alone, with absolutive and ergative cases. The order of morphemes in a transitive verb is fixed as [Aspect-B-A-Verb stem] (Koizumi et al., 2014). In Kaqchikel syntax, ergative case markers are called set A, and absolutive case markers are called set B. As an “Aspect” prefix, we used a *completive* marker “x- (pronounced [ʃ])” alone (similar to a suffix *-ed* or *-en* as a perfect participle in English). Each set further makes agreement of number (singular and plural) and person (first, second, and third) between a verb and object/subject (i.e., absolutive/ergative). In the present study, we used only *third* persons with the following prefixes:

- $\phi$ - (unmarked): singular for an absolutive case,
- e*-: plural for an absolutive case,
- r*-: singular for an ergative case, followed by a vowel-initial stem,
- ru*-: singular for an ergative case, followed by a consonant-initial stem,
- k*-: plural for an ergative case, followed by a vowel-initial stem, and

*ki*-: plural for an ergative case, followed by a consonant-initial stem.

## MATERIALS AND METHODS

### Participants

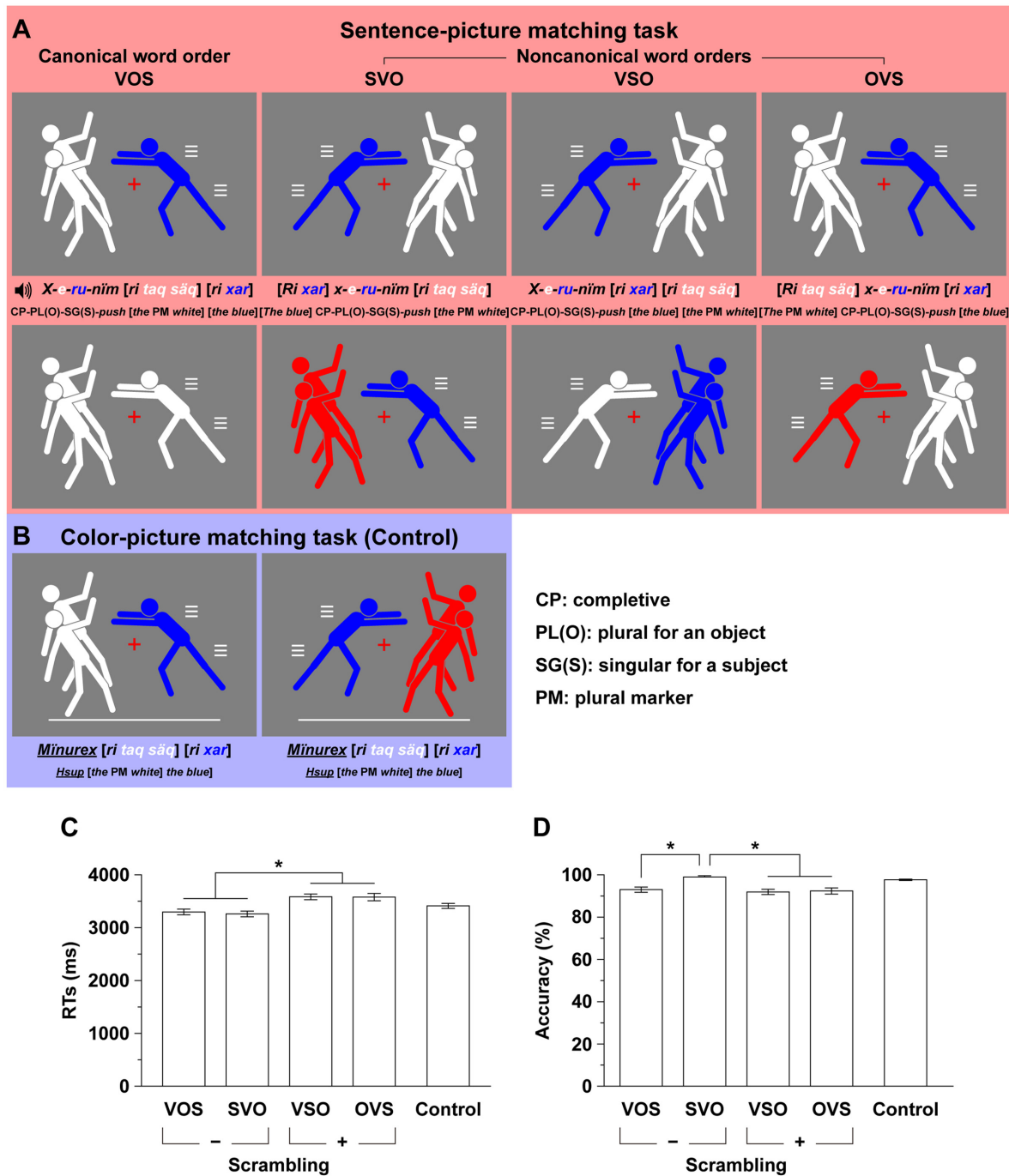
We recruited 20 Kaqchikel speakers, who lived in Chimaltenango, Sololá, or Sacatepéquez (the *Departamentos* of Guatemala). They spoke the Northern, Western, or Southern Kaqchikel dialects spread in these regions (six, nine, or two participants for each dialect, respectively). Recruiting Kaqchikel speakers for the present fMRI experiment was challenging, because they were not accustomed to being the participants of experiments and felt fatigue due to the unfamiliar environment in Tokyo during their week-long stay. One participant retired from the experiment after the second run. Two participants whose accuracy under the OVS condition was <75% were excluded from the subsequent behavioral and fMRI analyses. We eventually analyzed 17 participants (7 males, 10 females; mean  $\pm$  standard deviation [SD] age [years]: 32  $\pm$  7.9), who correctly achieved >75% correct answers under each of the four sentence conditions. This criterion was based on a model-based clustering analysis (Fraleigh and Raftery, 2002), in which the classification into two clusters showed the highest likelihood.

All of the 17 participants were Kaqchikel-Spanish bilinguals (age of acquisition of Spanish: 4.9  $\pm$  2.8 years), who showed right-handedness (laterality quotient: 72  $\pm$  28) as determined by the Edinburgh Handedness Inventory (Oldfield, 1971). None had a history of neurological disease. There were three Kaqchikel speakers from Patzún (a town in Guatemala) in the present study. A linguistic study of Kaqchikel reported that speakers in Patzún prefer a subject-initial word order in transitives and intransitives (e.g., SVO, SV), and that they tend to interpret the VOS word order as an interrogative sentence (Clemens, 2013). However, interrogative sentences can be clearly distinguished by a rise in intonation (García Matzar and Rodríguez Guaján, 1997), and there was no interrogative sentence in our stimuli.

Thirteen participants acquired Kaqchikel from infancy, and the other four participants acquired Kaqchikel from the age of 5–8 years (age of acquisition of Kaqchikel: 2.4  $\pm$  2.4 years). These four participants did not show any significant differences in the performance accuracy of the task compared to those who acquired Kaqchikel from infancy (two sample *t*-tests; [ $t(66) = 0.065$ ,  $p = 0.95$ ]). These four participants showed even shorter RTs [ $t(66) = 3.6$ ,  $p = 0.0007$ ], i.e., better performances. Moreover, a sub-analysis excluding these four participants showed basically similar activation patterns for the main effects of scrambling and topicalization.

During the experiments, translation was realized both ways through a Japanese-Spanish translator and a Spanish-Kaqchikel translator. To minimize the effects of Spanish usage during the experiment, we explained the stimuli and tasks to the participants in Kaqchikel through these translators. Prior to participation in





**FIGURE 2 | An experimental paradigm with various grammatical word orders in Kaqchikel. (A)** A sentence-picture matching task (marked in red). We tested four task conditions based on the different word orders: VOS, SVO, VSO, and OVS; the VOS is the canonical word order, and the others are *non-canonical* word orders that are always grammatical. Each sentence with one of these word orders is auditorily presented, and a simultaneously presented picture consisted of a single man and two men with the same or different colors (white, blue, red, or black). The participants judged whether a picture matched the meaning of the sentence. For each example sentence in Kaqchikel and its word-by-word translation in English, a pair of matched and mismatched pictures are shown in the first and second rows, respectively. For display purposes, the blue and white words match the blue and white men in the pictures of the first row, respectively. **(B)** A color-picture matching task (the control task; marked in blue). Examples of matched and mismatched stimuli are shown in the left and right panels of the third row, respectively. The participants judged whether the colors in a picture matched the color words in the auditory stimuli, irrespective of their order. **(C)** Reaction times (RTs) from the onset of the picture for the sentence-picture and control tasks. Only correct trials were included. Error bars indicate the standard error of the mean (SEM) for the participants. \*Corrected  $p < 0.05$ . **(D)** Accuracy for the sentence-picture and control tasks.

the study, written informed consent was obtained from each participant after the nature and possible consequences of the study were explained. Approval for the experiments was obtained from the institutional review board of the University of Tokyo, Komaba Campus.

## Stimuli

For each trial of the sentence-picture matching task, auditory and visual stimuli were simultaneously presented. As auditory stimuli, a set of 64 original sentences was prepared for matched stimuli (16 sentences for each sentence condition), and a set of 64 sentences, consisting of 36 original sentences and 28 additional sentences (6–8 for each condition), was used for mismatched stimuli (16 sentences for each condition). Here we call the sentence-picture stimuli *mismatched*, when a picture does not match the meaning of a sentence. All sentences were grammatical, and word frequencies were controlled among the conditions.

Under each of the four sentence conditions with VOS, SVO, VSO, and OVS, we used the same set of verbs, nouns, a definite article “*ri*,” and a plural marker for nouns “*taq*.” We used only men with a definite article for nouns, but did not use an indefinite article or an animal, because a mixed use of definiteness (definite and indefinite) or animacy (human being, animal, etc.) of noun phrases may affect word orders (García Matzar and Rodríguez Guaján, 1997). Either a single man or two men in a sentence were represented by one of four colors: “*käq* (*red*), *q’ëq* (*black*), *säq* (*white*), and *xar* (*blue*)” in Kaqchikel (see **Figure 2A**). We used one of the following six Kaqchikel verbs: “*chäy* (*hit*), *jik’* (*pull*), *nim* (*push*), *oyoj* (*call*), *pixab’aj* (*bless*), and *xib’ij* (*surprise*).” A sentence example with VOS is “*X-e-ru-nim [ri taq säq] [ri xar]* (*The blue pushed the whites*).”

In our stimuli, both the subject and object were humans. Note that the two sentences “*The blue pushed the white*” and “*The white pushed the blue*” (both men in singular or plural) cannot be distinguished by a prefix or noun phrase in Kaqchikel; the S and O cannot be formally determined. To resolve this type of ambiguity, each sentence included three men, which always consisted of one man (singular without “*taq*”) and two men (plural with “*taq*”).

All of the Kaqchikel sentences were spoken as a whole by a male native Kaqchikel speaker at a constant speed with natural prosody/intonation of declarative sentences, and those sentences were digitally recorded (16 bit; the normal audio cut-off, 44100 Hz). It should be noted that the spoken sentence contained rich information about prosody. With Sound Forge Pro 10 software (Sony Creative Software, Middleton, WI, USA), speech sounds were edited and their volumes were adjusted within the range from –50 to 0 dB full scale. A one-way repeated-measures analysis of variance (rANOVA) showed that the mean length of the auditory stimuli ( $2701 \pm 168$  ms) under each of the VOS, SVO, VSO, and OVS conditions was not significantly different [ $F(3,124) = 0.14, p = 0.94$ ]. The input volume was set to a comfortable hearing level for each participant.

As visual stimuli, a set of 16 original pictures was prepared for matched stimuli, which were used for every sentence condition. For mismatched stimuli, 64 pictures were additionally made (16 pictures for each sentence condition), in which either or

both of the color and number were changed from associated sentences. Half of the pictures depicted actions occurring from left to right, and the other half depicted actions occurring from right to left; colors of the single man and two men were also counterbalanced for both sides. The complexity of the pictures, as well as the frequency of action/color/number, was perfectly controlled among the sentence conditions.

All visual stimuli were presented against a gray background (**Figures 2A,B**). Each picture was presented for 5500 ms followed by a 500-ms blank interval. For fixation, a red cross was always shown at the center of the display, and the participants were instructed to keep their eyes on this position. Each auditory stimulus was presented 200 ms after the onset of each picture. The stimulus presentation and collection of behavioral data were controlled using the Presentation software package (Neurobehavioral Systems, Albany, CA, USA). The participants wore an MRI-compatible audio headset and an eyeglass-like MRI-compatible display (resolution,  $800 \times 600$ ; VisuaStim Digital, Resonance Technology, Northridge, CA, USA).

## The Sentence-Picture Matching and Color-Picture Matching Tasks

In the sentence-picture matching task, each participant listened to a Kaqchikel sentence and judged whether a picture matched the meaning of the sentence (**Figure 2A**). To minimize the inclusion of short term memory, we presented the sentence while the participant looked at the picture. Trials with matched and mismatched stimuli were presented equally often (16 trials each for matched and mismatched stimuli under one condition). They responded by pressing one of two buttons that were aligned in a row (right for the matched pair and left for the mismatched pair) with their right thumb. Matching a picture with a sentence required the four following linguistic properties:

- (1) color matching at the lowest lexical level,
- (2) plurality matching with or without the plural marker “*taq*,”
- (3) number and case (object/subject) matching, based on the verb prefixes, and
- (4) sentence construction based on syntactic structures.

The first property involved lexico-semantics, and for the next two properties, checking syntactic/semantic features was essential. For the last property, syntactic decisions were required. The judgment of mismatch was possible either at the phrase presented second or at the phrase presented third of the heard sentence, with the same frequency. Note that the comparison between trials with the matched and mismatched stimuli was not within the scope of the present study.

In the sentence-picture matching task, mismatched stimuli (e.g., pictures in the middle row of **Figure 2A**) involved only one of the following four variations: (1) 24 pictures (four or eight under each condition) with one color alone, while two colors were used in the sentence (e.g., the leftmost picture); (2) 16 pictures (four under each condition) with a color different from that used in the sentence (e.g., the second and fourth pictures), thereby controlling the frequency of colors under each condition; (3) eight pictures (four under the VOS and VSO conditions),

in which two colors were swapped between a single man and two men (e.g., the third picture); and (4) 16 pictures (eight each under the SVO and OVS conditions), in which the numbers of men were swapped. The first three variations of mismatched stimuli led to a violation in the linguistic properties mentioned above, thus requiring the comprehension of a whole sentence. The fourth variation, which involved attention to the exact verb prefixes due to the swapping of the number of men, may have required much higher loads than we had initially expected; we thus excluded those trials of mismatched stimuli from the subsequent behavioral and fMRI analyses.

In addition to the sentence-picture matching task, we also used a color-picture matching task (the *control* task), in which the participants judged whether colors in a picture matched the color words in the auditory stimuli (Figure 2B). By contrasting each of the four task conditions in the sentence-picture matching task with the control task at the first level of analysis, we could minimize the involvement of the first and second properties (see above) in any activation. For the auditory stimuli in the control task, we played the verb backward; as a result, the auditory stimuli contained the color words, plural marker, and definite articles. To indicate the control task, we added a white line at the bottom of the pictures, which were 128 different stimuli for the control task (64 each for matched and mismatched stimuli).

In the control task, mismatched stimuli involved only one of the following two variations: (1) 16 pictures with one color alone, while two colors were used in the auditory stimuli, and (2) 48 pictures with a color different from that in the auditory stimuli (e.g., the right picture of Figure 2B), thereby controlling the frequency of colors. General cognitive factors such as visual or auditory perception of the stimuli, matching, response selection, and motor responses were also controlled by the control task. We used the control condition as a baseline of the first-level analyses of the fMRI data to exclude these sensory and general cognitive factors as much as possible. The participants underwent short practice sessions before the task sessions to become fully familiarized with these tasks.

A single run of the task sessions (192 s) contained 16 “test events” of the sentence-picture matching task (four times each for the VOS, SVO, VSO, and OVS conditions), with inter-trial intervals of one control task. The order of the test events was pseudorandomized without repetition of the same condition, to prevent any condition-specific strategy. A single run contained 16 trials of the control task. Seven or eight runs were tested per one participant in a day. Only trials with participants’ correct responses were used for analyzing the RTs and fMRI data. For each participant, seven or eight runs without head movement were used for the behavior and fMRI analyses.

## MRI Data Acquisition

For the MRI data acquisition, the participant was in a supine position, and his or her head was immobilized inside the radio-frequency coil with straps. The MRI scans were conducted on a 3.0 T MRI system (GE Signa HDxt 3.0T; GE Healthcare, Milwaukee, WI, USA). We scanned 32 axial slices of 3-mm thick with a 0.3-mm gap, covering the volume range of  $-42.9$  to  $62.4$  mm from the anterior to posterior commissure (AC-PC)

line in the vertical direction, using a gradient-echo echo-planar imaging (EPI) sequence (repetition time [TR] = 2 s, echo time [TE] = 30 ms, flip angle [FA] =  $90^\circ$ , field of view [FOV] =  $192$  mm  $\times$   $192$  mm, resolution =  $3$  mm  $\times$   $3$  mm). In a single scan, we obtained 102 volumes where the first six images were discarded, which allowed for the rise of the MR signals.

After the completion of the fMRI session, high-resolution T1-weighted images of the whole brain (192 axial slices,  $1.0$  mm  $\times$   $1.0$  mm  $\times$   $1.0$  mm) were acquired from all participants with a three-dimensional fast spoiled gradient recalled acquisition in the steady state (3D FSPGR) sequence (TR = 8.6 ms, TE = 2.6 ms, FA =  $25^\circ$ , FOV =  $256$  mm  $\times$   $256$  mm). These structural images were used for normalizing the fMRI data.

## fMRI Data Analyses

The fMRI data were analyzed in a standard manner using SPM12 statistical parametric mapping software (Wellcome Trust Centre for Neuroimaging<sup>1</sup>) (Friston et al., 1995), implemented on MATLAB software (MathWorks, Natick, MA, USA). The acquisition timing of each slice was corrected using the middle slice (the 17th slice chronologically) as a reference for the EPI data. We realigned the time-series data in multiple runs to the first volume in all runs, and further realigned the data to the mean volume of all runs. The realigned data were resliced using seventh-degree B-spline interpolation, so that each voxel of each functional image matched that of the first volume. We removed runs that included data with a translation of  $>2$  mm in any of the three directions and with a rotation of  $>1.4^\circ$  around any of the three axes; these thresholds of head movement were empirically determined from our previous studies (Hashimoto and Sakai, 2002; Suzuki and Sakai, 2003; Kinno et al., 2008; Ohta et al., 2013b). For this reason, a single run was removed from one participant.

After alignment to the AC-PC line, each participant’s T1-weighted structural image was coregistered to the mean functional image generated during realignment. T1-weighted images were bias-corrected with light regularization, and segmented to the gray matter, white matter, cerebrospinal fluid, bone, other soft tissues, and air by using default tissue probability maps and the Segment tool in the SPM12, which uses an affine regularization to warp images to the International Consortium for Brain Mapping European brain template (Ashburner and Friston, 2005). Inter-subject registration was achieved with Diffeomorphic Anatomical Registration using the Exponentiated Lie algebra (DARTEL) toolbox in the SPM12 (Ashburner, 2007). The coregistered structural images were spatially normalized to the standard brain space as defined by the Montreal Neurological Institute (MNI) using DARTEL’s Normalize to MNI Space tool. All of the normalized structural images were visually inspected and compared with the standard brain for the absence of any further deformation. The realigned functional images were also spatially normalized to the MNI space by using DARTEL’s Normalize to MNI Space tool, which converted voxel sizes to

<sup>1</sup><http://www.fil.ion.ucl.ac.uk/spm/>

3 mm × 3 mm × 3 mm and smoothed the images with an isotropic Gaussian kernel of 9-mm full-width at half maximum.

In a first-level analysis (i.e., the fixed-effects analysis), each participant's hemodynamic responses induced by the four sentence conditions as well as the control task for each session were modeled with a boxcar function with a duration of 5.5 s from the onset of each visual stimulus. The boxcar function was then convolved with a hemodynamic response function. Low-frequency noise was removed by high-pass filtering at 1/128 Hz. To minimize the effects of head movement, the six realignment parameters obtained from preprocessing were included as a nuisance factor in a general linear model. The images of the VOS – control, SVO – control, VSO – control, and OVS – control contrasts were then generated in the general linear model for each participant and used for the intersubject comparison in a second-level analysis (i.e., the random-effects analysis). To examine the activation of the regions in an unbiased manner, we adopted whole-brain analyses (Friston and Henson, 2006).

A repeated-measures analysis of covariance with *t*-tests was performed with two factors (scrambling × topicalization), the results of which were thresholded at uncorrected  $p < 0.0001$  ( $t > 4.8$ ) for the voxel level, and at corrected  $p < 0.05$  for the cluster level, with topological false discovery rate (FDR) correction across the whole brain (Chumbley and Friston, 2009). We used the differences of accuracy between each sentence condition and control (e.g., VOS – control, SVO – control, VSO – control, and OVS – control) as a covariate of no interest (i.e., a nuisance factor). For the anatomical identification of activated regions, we basically used the Anatomical Automatic Labeling method<sup>2</sup> (Tzourio-Mazoyer et al., 2002) and the labeled data as provided by Neuromorphometrics Inc.<sup>3</sup> under academic subscription. For each region of interest, we extracted the mean percent signal changes for each participant from the local maxima (i.e., peak voxel) of each region in the second-level group analysis, using the MarsBaR-toolbox<sup>4</sup>.

## RESULTS

### Behavioral Data

We used a two-by-two experimental design (factors: scrambling × topicalization). The behavioral data for the sentence-picture matching task are shown in **Figures 2C,D**. Under the sentence conditions, an rANOVA with these two factors on the RTs showed significant main effects of scrambling [ $F(1,16) = 153$ ,  $p < 0.0001$ ], but the main effect of topicalization and an interaction between these factors were not significant [topicalization,  $F(1,16) = 0.61$ ,  $p = 0.45$ ; interaction,  $F(1,16) = 0.24$ ,  $p = 0.63$ ]. Consistent with our theoretical predictions, these results indicated that the VSO and OVS conditions [+scrambling] produce greater syntactic loads than the VOS and SVO conditions [–scrambling].

Regarding the accuracy, the participants made reliable and consistent judgments, and the accuracy under every condition was higher than 90%. Under the sentence conditions, an rANOVA with these two factors on the accuracy showed significant main effects of scrambling and topicalization [scrambling,  $F(1,16) = 12$ ,  $p = 0.0036$ ; topicalization,  $F(1,16) = 9.5$ ,  $p = 0.0072$ ], and an interaction between these factors was marginally significant [ $F(1,16) = 4.4$ ,  $p = 0.052$ ]. *Post hoc* paired *t*-tests showed that the accuracy under the SVO condition was significantly higher than that under the other conditions (corrected  $p < 0.0024$ ), indicating that SVO was the easiest condition.

### The Basic Design of the Functional Analyses

Here we outline the basic design of the main functional analyses. Based on the two-by-two experimental design (scrambling × topicalization), we first examined the main effects of scrambling [ $\pm S$ ], i.e., (VSO + OVS) – (VOS + SVO), where the [+S] conditions mainly induced higher syntactic loads (see the Introduction). We then examined the main effects of topicalization [ $\pm T$ ], i.e., (VOS + VSO) vs. (SVO + OVS), related to the semantic/phonological loads. To examine any effects associated with the accuracy for each condition, we also tested (VOS + VSO + OVS) – SVO, based on the behavioral results shown above.

To exclude any semantic/phonological effects of the object-subject word orders, we performed two direct comparisons while making the factor of topicalization constant: VSO [+S, –T] vs. VOS [–S, –T], and OVS [+S, +T] vs. SVO [–S, +T]. Lastly, we examined the activation profiles under the four sentence conditions in each of the identified regions of interest, and the results confirmed significant activation in these regions for a diagonal contrast of VSO [+S, –T] vs. SVO [–S, +T].

### The Cortical Activation Reflecting Syntactic Loads or Semantic/Phonological Loads

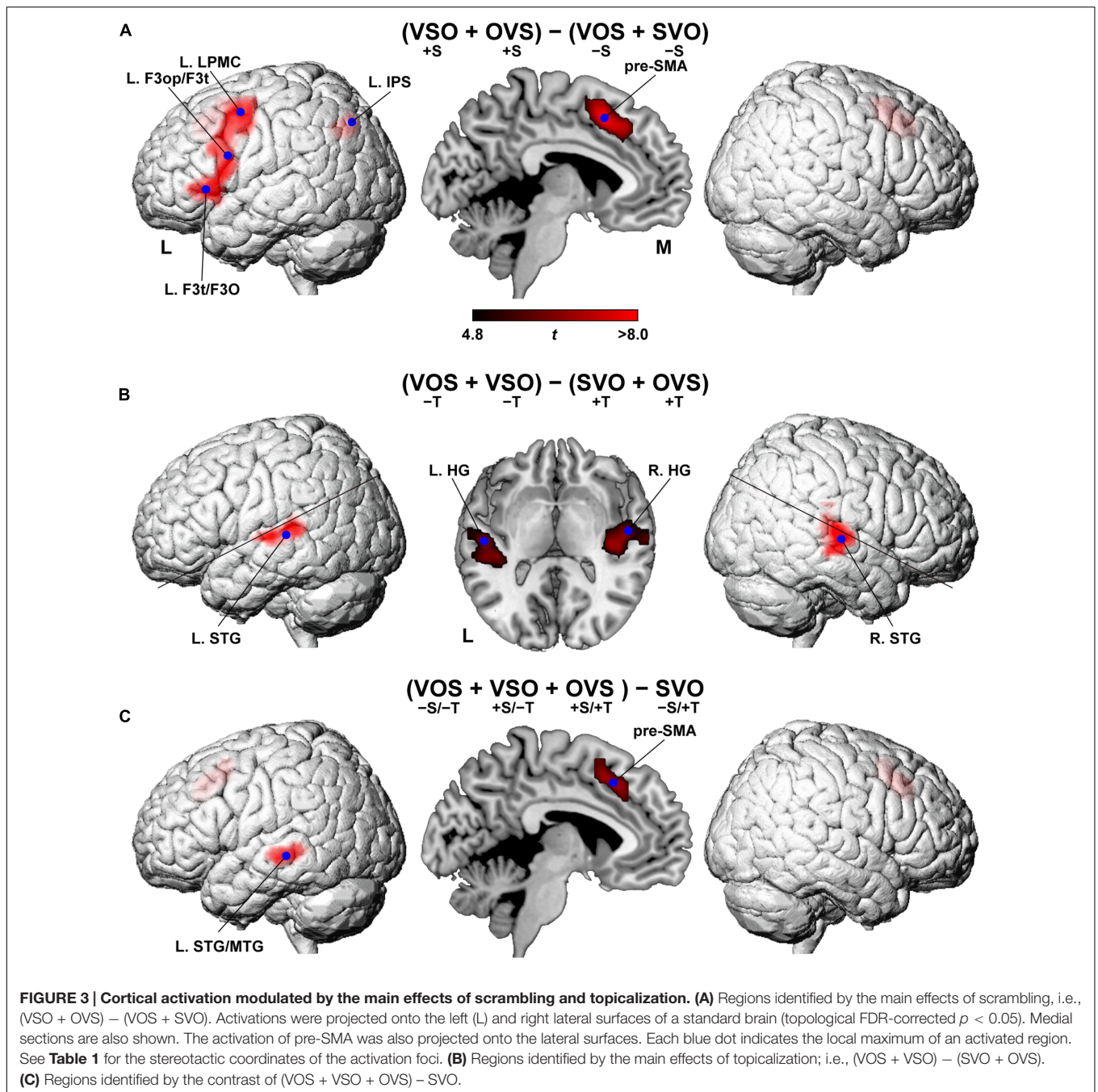
The main effects of scrambling, i.e., (VSO + OVS) – (VOS + SVO), were observed in language areas such as the L. LPMC, L. F3op/F3t, and L. F3t/F3O (corrected  $p < 0.05$ ) (**Figure 3A** and **Table 1**). Additional activation was observed in the pre-supplementary motor area (pre-SMA) and the left intraparietal sulcus (L. IPS). In contrast, the main effects of topicalization, i.e., (VOS + VSO) – (SVO + OVS), were observed in completely different regions: Heschl's gyrus (HG) and the STG in both hemispheres (**Figure 3B** and **Table 1**). The reverse contrast, i.e., (SVO + OVS) – (VOS + VSO), did not show any significant activation (corrected  $p > 0.9$ ). These results support the possibility that phonological loads and general auditory attention would become larger in the absence of topicalization (see the Introduction).

In contrast, the contrast of (VOS + VSO + OVS) – SVO showed activation in the pre-SMA and left superior and middle temporal gyri (L. STG/MTG), sparing the lateral frontal regions (**Figure 3C** and **Table 1**). The pre-SMA activation

<sup>2</sup><http://www.gin.cnrs.fr/AAL2/>

<sup>3</sup><http://Neuromorphometrics.com/>

<sup>4</sup><http://marsbar.sourceforge.net/>



replicated activation in the main effects of scrambling, while the L. STG/MTG activation was left-lateralized and located more ventrally than that in the main effects of topicalization.

We directly compared the cortical activation in  $VSO - VOS$ , and we observed localized activation in the L. LPMC, L. F3op/F3t, and L. F3t/F30 (**Figure 4A** and **Table 1**), i.e., the frontal language areas, which were consistent with the main effects of scrambling. Activation in the pre-SMA and L. IPS also replicated the main effects of scrambling, but the R. IPS was additionally activated. On the other hand, the reverse contrast, i.e.,  $VOS - VSO$ , did not show any significant activation (corrected  $p > 0.9$ ). In  $OVS -$

$SVO$ , the overall activation pattern was similar to that in  $VSO - VOS$  (**Figure 4B** and **Table 1**). It is notable that the L. F3op/F3t activation shifted more dorsally (15 mm for the local maxima) in  $OVS - SVO$ . Compared with the main effects of scrambling, these activated regions were highly localized in such stringent contrasts as  $VSO - VOS$  and  $OVS - SVO$ .

At the local maxima of the L. LPMC, L. F3op/F3t, L. F3t/F30, and pre-SMA in the second-level analysis, which were selected from the contrast  $OVS - SVO$  (shown as blue dots in **Figure 4B**), we examined the percent signal changes. In all of these regions, the overall activation profiles under the four sentence conditions

**TABLE 1 | Regions identified by the effects of word order.**

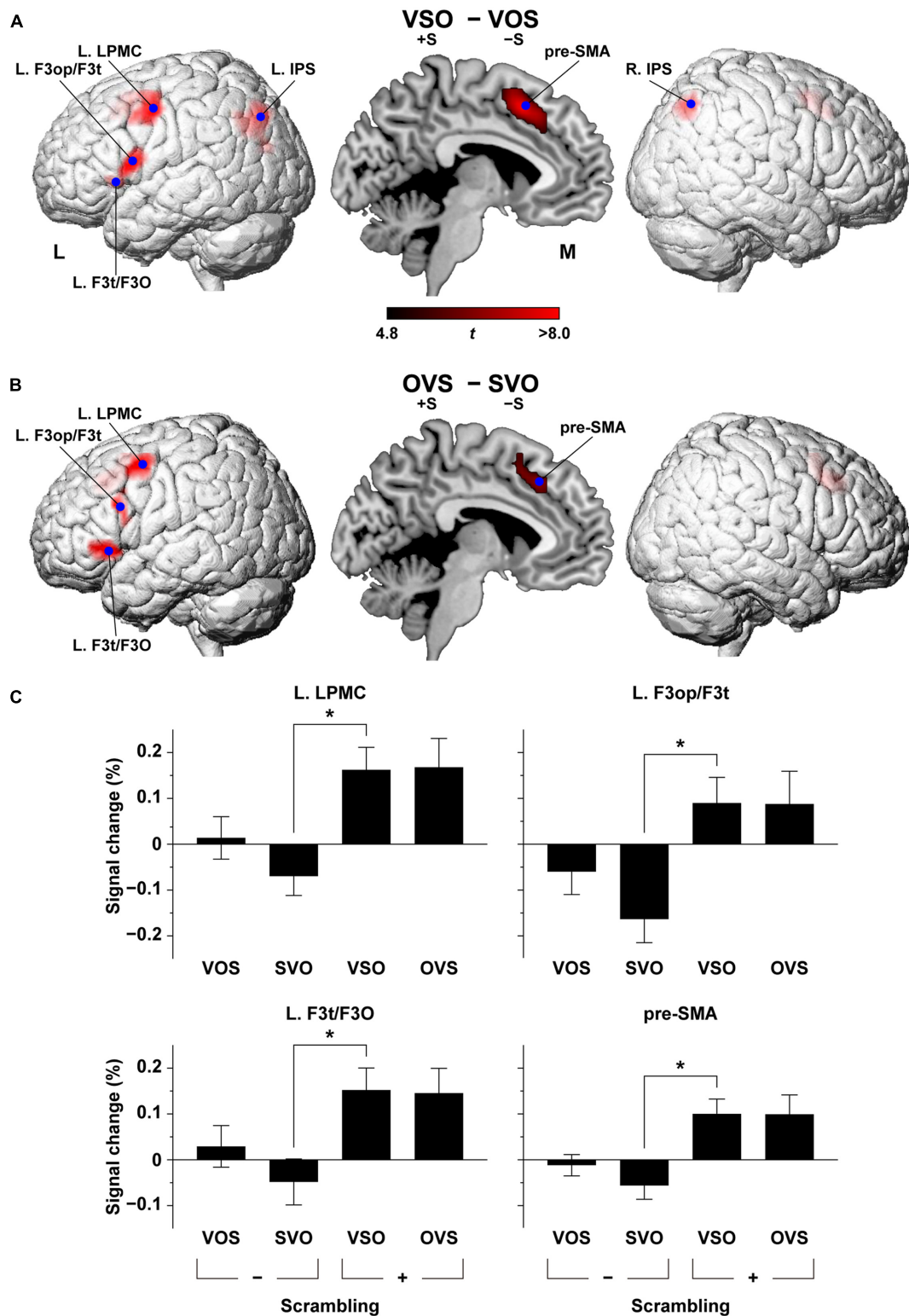
Brain regions	BA	Side	x	y	z	Z	Voxels
<b>Main effects of scrambling: (VSO + OVS) – (VOS + SVO)</b>							
LPMC	6/9	L	–42	3	51	6.3	625
			–45	12	42	5.4	*
			–36	12	33	4.9	*
F3op/F3t	44/45	L	–48	18	27	5.1	*
			–54	12	15	5.8	*
F3t/F3O	45/47	L	–54	27	0	6.1	*
pre-SMA	6/8/32	M	–6	15	48	6.5	352
			0	24	42	6.2	*
IPS	7	L	–27	–69	42	5.2	71
<b>Main effects of topicalization: (VOS + VSO) – (SVO + OVS)</b>							
HG	41/42	L	–51	–21	6	5.6	213
STG	22	L	–51	–33	9	5.7	*
			–63	–15	3	4.7	*
HG	41/42	R	48	–15	6	5.6	326
			39	–24	6	5.8	*
STG	22	R	66	–18	9	5.6	*
			63	–18	–6	5.1	*
<b>(VOS + VSO + OVS) – SVO</b>							
pre-SMA	6/8/32	M	–6	15	48	5.5	235
			0	27	42	5.7	*
STG/MTG	22/21	L	–63	–21	–6	5.7	96
<b>VSO – VOS</b>							
LPMC	6/8	L	–36	6	57	4.5	112
			–48	0	45	5.4	*
			–48	12	42	4.6	*
F3op/F3t	44/45	L	–57	12	12	5.5	78
F3t/F3O	45/47	L	–54	24	0	4.4	*
pre-SMA	6/8/32	M	–6	15	48	5.4	134
IPS	7/19	L	–12	–78	45	4.7	149
			–27	–69	39	5.5	*
			–30	–75	24	4.7	*
	7/19	R	18	–69	51	5.5	95
			30	–66	51	5.1	*
<b>OVS – SVO</b>							
LPMC	6/8	L	–42	3	54	5.5	74
F3op/F3t	44/45	L	–51	18	27	4.7	35
F3t/F3O	45/47	L	–54	30	0	5.5	46
pre-SMA	6/8/32	M	–3	15	60	4.9	149
			–6	15	48	5.1	*
			–3	27	42	5.5	*

Stereotactic coordinates ( $x$ ,  $y$ ,  $z$ ) in the MNI space (mm) are shown for each activation peak of  $Z$  values. The threshold was set at corrected  $p < 0.05$  for the cluster level. BA = Brodmann's area; L = left; M = medial; R = right; LPMC = lateral premotor cortex; F3op/F3t/F3O = opercular/triangular/orbital parts of the inferior frontal gyrus; pre-SMA = pre-supplementary motor area; IPS = intraparietal sulcus; HG = Heschl's gyrus; STG = superior temporal gyrus; MTG = middle temporal gyrus. The region with an asterisk is included within the same cluster shown one row above.

were consistent. More specifically, the activations under VSO and OVS were always evident at the same level, whereas the activations under VOS and SVO were near or below the baseline level. Furthermore, the signal changes under VSO were significantly larger than those under SVO in each of the four regions (corrected  $p < 0.002$ ).

The bilateral IPS activation, which was observed in VSO – VOS (shown as blue dots in **Figure 4A**), but not in

OVS – SVO, may indicate the presence of an interaction. This effect was due to more activations under the VSO condition than the other conditions. A significant interaction was present in the R. IPS [ $F(1,16) = 6.5$ ,  $p = 0.022$ ], but not in the L. IPS [ $F(1,16) = 0.17$ ,  $p = 0.69$ ]. The VSO [+S, –T] condition reflected a synergistic effect of multiple linguistic factors, which may employ additional cortical regions like the bilateral IPS.



**FIGURE 4 | Direct comparison of cortical activation between conditions.** The VSO – VOS contrast (A) and the OVS – SVO contrast (B). Activations were projected onto the left (L) and right lateral surfaces of a standard brain (topological FDR-corrected  $p < 0.05$ ). See Table 1 for the stereotactic coordinates of the activation foci. (C) Histograms for the percent signal changes at the local maxima of the L. LPMC, L. F3op/F3t, L. F3t/F3O, and pre-SMA in OVS – SVO. The signal changes for VOS, SVO, VSO, and OVS are shown with reference to the baseline level of the control task. Error bars indicate SEM for the participants.  $*p < 0.0005$ .

## DISCUSSION

By using the sentence-picture matching task in the Kaqchikel language, we obtained four striking results. First, we found that the [+scrambling] conditions elicited significant activation in the left frontal regions of the L. LPMC, L. F3op/F3t, and L. F3t/F3O (Figure 3A), indicating the effects of syntactic loads in Kaqchikel, a head-marking and OS language. These results indicate that the L. LPMC, L. F3op/F3t, and L. F3t/F3O, but not the L. ATL, are crucial for a movement of phrases. Secondly, the [–topicalization] conditions resulted in significant activation in the bilateral HG and STG (Figure 3B), demonstrating that the syntactic and phonological processes were clearly dissociated within the language areas. Thirdly, the pre-SMA and L. STG/MTG were activated under the more demanding conditions other than SVO (Figure 3C), suggesting their supportive roles in syntactic or semantic processing. Fourthly, two direct comparisons of VSO – VOS and OVS – SVO showed *consistent* and localized activations in the L. LPMC, L. F3op/F3t, and L. F3t/F3O, as well as the pre-SMA (Figures 4A,B), while VOS – VSO did not show any significant activation. This last point fits the syntactic account for the selective activation in these frontal regions, excluding any semantic effects of the OS word order itself, which might be related to “irregular prominence factors of noun phrases” (see the Introduction). Our findings further indicate that the functional roles of these left frontal and temporal regions involve *linguistic* aspects themselves, namely syntax versus semantics/phonology, rather than output/input aspects of speech processing. Moreover, the present study with Kaqchikel clearly contributes to the concept that such universal operations as scrambling and topicalization are differentially processed in specified cortical regions.

“Merge” is a fundamental local structure-building operation proposed by modern linguistics (Chomsky, 1995), and is a key to syntactic processing. Neuroimaging studies have established that syntactic processing selectively activates the L. F3op/F3t and L. LPMC (Stromswold et al., 1996; Dapretto and Bookheimer, 1999; Embick et al., 2000; Hashimoto and Sakai, 2002; Friederici et al., 2003; Musso et al., 2003), indicating that these regions have a critical role as grammar centers (Sakai, 2005). Activations in the L. F3op/F3t and L. LPMC have also been observed in our studies using Japanese sentences with non-canonical word orders (Kinno et al., 2008). Moreover, our MEG studies showed a significant increase of responses in the L. IFG, which reflected predictive effects on a verb caused by a preceding object in a short sentence (Iijima et al., 2009; Inubushi et al., 2012; Iijima and Sakai, 2014). In the present study, we observed selective activation in the L. F3op/F3t and L. LPMC under the [+scrambling] conditions, which is consistent with these previous findings. Our results also support the explanation based on the DoM (Ohta et al., 2013a,b), in that the [+scrambling] conditions with the larger DoM enhanced the L. F3op/F3t and L. LPMC activations. It should be noted that activation in the L. LPMC, L. F3op/F3t, and L. F3t/F3O were more localized in both VSO – VOS and OVS – SVO, which excluded any differences in semantic/phonological loads. To our knowledge, our present findings are the first

experimental evidence of linguistic computation that dissociates [+scrambling] and [–topicalization].

Here we observed activation in the bilateral HG and STG under the [–topicalization] conditions, which may reflect phonological loads and attention in the absence of topicalization. Our previous fMRI study revealed that the bilateral STG activations were selectively enhanced by phonological decision tasks (Suzuki and Sakai, 2003). The same study further demonstrated that the localized activations in the L. MTG were modulated by the presence of syntactic or semantic errors, which may enhance processing loads to correct sentences. Consistent with this possibility, here we observed the localized L. STG/MTG activation in the contrast of (VOS + VSO + OVS) – SVO.

In recent studies using a visual sentence-picture matching task similar to that used here, we tested 21 patients with a left frontal glioma and observed abnormal overactivity and/or underactivity in 14 syntax-related regions (Kinno et al., 2014, 2015). Those investigations also revealed three syntax-related networks: network I (syntax and its supportive system), network II (syntax and input/output interface), and network III (syntax and semantics). Functional and anatomical connectivity was observed within individual networks in normal controls, whereas in the agrammatic patients almost all of the functional connectivity exhibited chaotic changes. Moreover, the patients who showed normal performances showed normal connectivity between the L. F3op/F3t and L. IPS, as well as normal connectivity between the L. F3t and L. F3O, indicating that these pathways are the most crucial among the syntax-related networks. In the present study, we observed significant activation in the pre-SMA and L. IPS (Figures 3, 4), which are included in network I (which consists of the L. F3op/F3t, pre-SMA, right lateral frontal regions, L. IPS, and right temporal regions). The consistent activation of the pre-SMA and L. IPS suggests their supportive roles in syntactic processing.

Another possible model for the syntactic structures of Mayan sentences has been proposed in a linguistic study: a predicate-fronting model (Figure 1B; Coon, 2010). In this model, which is more complex than the right-specifier model even for the canonical word order (VOS), *predicate fronting* is assumed as a default and obligatory movement. This model was based on similar syntactic analyses for another verb-initial language: Niuean, a Polynesian language that is markedly distant from Mayan languages (Massam, 2010). The notion of *object shift*, which precedes predicate fronting, replaces scrambling in the right-specifier model. Note that the verification of the right-specifier model or predicate fronting model was not within the scope of the present study; both models predict consistent results in our paradigm. The explanatory adequacy of these two models should be further examined in future experiments.

A previous fMRI study in Kaqchikel with a sentence plausibility judgment task has reported higher activation in the left IFG close to its border with the left middle frontal gyrus (BA 46) in the SVO – VOS contrast (Koizumi and Kim, 2016), clearly different activation from the present results. In that task, the participants listened to a sentence with a human (S) and an inanimate entity (O), and judged whether a sentence was semantically plausible or not, where no specific



context was provided. The authors of that study interpreted this activation as the higher processing load related to more complex structures of SVO. They further argued that this higher load was related to the discourse-pragmatic requirements for the non-canonical SVO word order. A topicalized sentence incurs higher processing loads when presented out of context, as in the case of the sentence plausibility judgment task. In our present study, where both the subject and object were humans, we observed significant activation in the bilateral HG and STG as the main effects of topicalization, i.e., (VOS + VSO) – (SVO + OVS). This activation reflected increased phonological loads under the [–topicalization] conditions. Note that the reverse contrast did not show any significant activation, indicating that the [+topicalization] conditions had little syntactic effects where DoM remains minimal (see the Introduction). By naturally providing a discourse context as a picture, we were able to dissociate the main effects of topicalization related to semantics/phonology from those related to pragmatics.

In Kaqchikel, it has been reported that SVO is more frequently used than VOS (73% versus 15%) (Kubo et al., 2011). In that study, the native Kaqchikel speakers made a sentence describing a picture, which depicted a transitive action between a human agent and human/animal/inanimate patient. Although the concept of “basic word order” has been problematic (Brody, 1984), the word order with the simplest syntactic structure, i.e., syntactically canonical word order, is VOS (García Matzar and Rodríguez Guaján, 1997, p. 333). It has been suggested that “when examining the *basic* word order of Mayan languages, *syntactically* determined word order from the standpoint of syntactic complexity needs to be distinguished from *pragmatically* determined word order, commonly used for pragmatic purposes” (Yasunaga et al., 2015). This point is also related to our present observation that SVO was the easiest condition in our paradigm (Figure 2D). Both the higher production frequency and the higher accuracy of SVO may be caused by the effects associated with [+topicalization]. In a study using Japanese sentences, the production frequency of subject-topicalized sentences (S-*wa* OV) was several times higher than that of canonical sentences (S-*ga* OV), and the subject-topicalized sentences were more easily processed than the object-topicalized sentences (O-*wa* S-*ga* V) (Imamura and Koizumi, 2011); note

that an S is considered as a default topic (Koster, 1978). These phenomena could be parallel to those for SVO versus canonical VOS in Kaqchikel, in that subject-topicalized sentences have lower semantic/phonological loads. On the other hand, one crucial difference between Japanese and Kaqchikel is that both canonical and subject-topicalized sentences are SOV in Japanese. Moreover, a string-vacuous movement, i.e., a movement without a change in the order of strings, is prohibited (Chomsky, 1986). Because both scrambling and topicalization are not string-vacuous movements in Kaqchikel, Kaqchikel is an ideal language for dissociating the effects of scrambling and topicalization. By targeting such understudied languages as Kaqchikel, we were able to integrate previous findings of neuroimaging and linguistic studies with our new findings, which will contribute to the understanding of the biological basis of language.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This research was supported by a Core Research for Evolutional Science and Technology (CREST) grant from the Japan Science and Technology Agency (JST), a Grant-in-Aid for Scientific Research (A; No. 15H02603), a Grant-in-Aid for Young Scientists (B; No. 15K16733) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and a Grant-in-Aid for Japan Society for the Promotion of Science (JSPS) Fellows (No. 24-8931).

## ACKNOWLEDGMENTS

We thank J. E. Ajsivinac Sian, L. P. O. García Matzar, and A. Yamagata (a Spanish interpreter) for their assistance with the experiments, N. Komoro and H. Tada for their technical assistance, and H. Matsuda for her administrative assistance.

## REFERENCES

- Aissen, J. L. (1992). Topic and focus in Mayan. *Language* 68, 43–80. doi: 10.2307/416369
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage* 38, 95–113. doi: 10.1016/j.neuroimage.2007.07.007
- Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *Neuroimage* 26, 839–851. doi: 10.1016/j.neuroimage.2005.02.018
- Bahlmann, J., Rodríguez-Fornells, A., Rotte, M., and Münte, T. F. (2007). An fMRI study of canonical and noncanonical word order in German. *Hum. Brain Mapp.* 28, 940–949. doi: 10.1002/hbm.20318
- Ben-Shachar, M., Palti, D., and Grodzinsky, Y. (2004). Neural correlates of syntactic movement: converging evidence from two fMRI experiments. *Neuroimage* 21, 1320–1336. doi: 10.1016/j.neuroimage.2003.11.027
- Bornkessel, I., Zysset, S., Friederici, A. D., Von Cramon, D. Y., and Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *Neuroimage* 26, 221–233. doi: 10.1016/j.neuroimage.2005.01.032
- Brody, J. (1984). Some problems with the concept of basic word order. *Linguistics* 22, 711–736. doi: 10.1515/ling.1984.22.5.711
- Brown, R. M., Maxwell, J. M., and Little, W. E. (2006). *La üt̕z awäch?: Introduction to Kaqchikel Maya Language*. Austin, TX: University of Texas Press.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton Publishers.
- Chomsky, N. (1986). *Barriers*. Cambridge, MA: The MIT Press.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: The MIT Press.
- Chumbley, J. R., and Friston, K. J. (2009). False discovery rate revisited: FDR and topological inference using Gaussian random fields. *Neuroimage* 44, 62–70. doi: 10.1016/j.neuroimage.2008.05.021
- Clemens, L. E. (2013). “Kaqchikel SVO: V2 in a V1 language,” in *Studies in Kaqchikel Grammar*, ed. M. Kenstowicz (Cambridge, MA: MITWPL), 1–23.
- Coon, J. (2010). VOS as predicate fronting in Chol. *Lingua* 120, 354–378. doi: 10.1016/j.lingua.2008.07.006

- Dapretto, M., and Bookheimer, S. Y. (1999). Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron* 24, 427–432. doi: 10.1016/s0896-6273(00)80855-7
- Embick, D., Marantz, A., Miyashita, Y., O'neil, W., and Sakai, K. L. (2000). A syntactic specialization for Broca's area. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6150–6154. doi: 10.1073/pnas.100098897
- Felser, C., Clahsen, H., and Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: an ERP study of topicalization and *wh*-movement in German. *Brain Lang.* 87, 345–354. doi: 10.1016/S0093-934X(03)00135-4
- Fraley, C., and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* 97, 611–631. doi: 10.1198/016214502760047131
- Friederici, A. D., Rüschemeyer, S.-A., Hahne, A., and Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb. Cortex* 13, 170–177. doi: 10.1093/cercor/13.2.170
- Friston, K. J., and Henson, R. N. (2006). Commentary on: divide and conquer; a defence of functional localisers. *Neuroimage* 30, 1097–1099. doi: 10.1016/j.neuroimage.2006.02.007
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210. doi: 10.1002/hbm.460020402
- Fukui, N. (1993). Parameter and optionality. *Linguist. Inq.* 24, 399–420.
- García Matzar, L. P., and Rodríguez Guaján, P. B. J. O. (1997). *Rukemik ri Kaqchikel chi': Gramática Kaqchikel [Kaqchikel Grammar]*. Guatemala City: Cholsamaj.
- Grewe, T., Bornkessel, I., Zysset, S., Wiese, R., Von Cramon, D. Y., and Schleuwsky, M. (2006). Linguistic prominence and Broca's area: the influence of animacy as a linearization principle. *Neuroimage* 32, 1395–1402. doi: 10.1016/j.neuroimage.2006.04.213
- Hashimoto, R., and Sakai, K. L. (2002). Specialization in the left prefrontal cortex for sentence comprehension. *Neuron* 35, 589–597. doi: 10.1016/s0896-6273(02)00788-2
- Iijima, K., Fukui, N., and Sakai, K. L. (2009). The cortical dynamics in building syntactic structures of sentences: an MEG study in a minimal-pair paradigm. *Neuroimage* 44, 1387–1396. doi: 10.1016/j.neuroimage.2008.10.041
- Iijima, K., and Sakai, K. L. (2014). Subliminal enhancement of predictive effects during syntactic processing in the left inferior frontal gyrus: an MEG study. *Front. Syst. Neurosci.* 8:217. doi: 10.3389/fnsys.2014.00217
- Imamura, S., and Koizumi, M. (2011). A centering analysis of word order in Japanese. *Tohoku Stud. Linguist.* 20, 59–74.
- Inubushi, T., Iijima, K., Koizumi, M., and Sakai, K. L. (2012). Left inferior frontal activations depending on the canonicity determined by the argument structures of ditransitive sentences: an MEG study. *PLoS ONE* 7:e37192. doi: 10.1371/journal.pone.0037192
- Karimi, S. (ed.). (2003). *Word Order and Scrambling*. Malden, MA: Blackwell Publishing.
- Kim, J., Koizumi, M., Ikuta, N., Fukumitsu, Y., Kimura, N., Iwata, K., et al. (2009). Scrambling effects on the processing of Japanese sentences: an fMRI study. *J. Neurolinguist.* 22, 151–166. doi: 10.1016/j.jneuroling.2008.07.005
- Kinno, R., Kawamura, M., Shioda, S., and Sakai, K. L. (2008). Neural correlates of noncanonical syntactic processing revealed by a picture-sentence matching task. *Hum. Brain Mapp.* 29, 1015–1027. doi: 10.1002/hbm.20441
- Kinno, R., Ohta, S., Muragaki, Y., Maruyama, T., and Sakai, K. L. (2014). Differential reorganization of three syntax-related networks induced by a left frontal glioma. *Brain* 137, 1193–1212. doi: 10.1093/brain/awu013
- Kinno, R., Ohta, S., Muragaki, Y., Maruyama, T., and Sakai, K. L. (2015). Left frontal glioma induces functional connectivity changes in syntax-related networks. *Springerplus* 4:317. doi: 10.1186/s40064-015-1104-6
- Koizumi, M., and Kim, J. (2016). Greater left inferior frontal activation for SVO than VOS during sentence comprehension in Kaqchikel. *Front. Psychol.* 7:1541. doi: 10.3389/fpsyg.2016.01541
- Koizumi, M., Yasugi, Y., Tamaoka, K., Kiyama, S., Kim, J., Ajsivnac Sian, J. E., et al. (2014). On the (non)universality of the preference for subject-object word order in sentence comprehension: a sentence-processing study in Kaqchikel Maya. *Language* 90, 722–736. doi: 10.1353/lan.2014.0068
- Koster, J. (1978). "Why subject sentences don't exist," in *Recent Transformational Studies in European Languages*, ed. S. J. Keyser (Cambridge, MA: MIT Press), 53–64.
- Kubo, T., Ono, H., Tanaka, M., Koizumi, M., and Sakai, H. (2011). VOS-gengo ni oite yuseisei ga gojun ni ataeru eikyo: kakuchikerugo ni okeru sengabyosha kadai de no kento [Animacy effects on word order in VOS language: evidence from a picture description task in Kaqchikel]. *IEICE Tech. Rep.* 111, 19–24.
- Marantz, A. (2005). Generative linguistics within the cognitive neuroscience of language. *Linguist. Rev.* 22, 429–445. doi: 10.1515/tlr.2005.22.2-4.429
- Massam, D. (2010). V1 or V2?: on the left in Niuean. *Lingua* 120, 284–302. doi: 10.1016/j.lingua.2008.11.014
- Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., et al. (2003). Broca's area and the language instinct. *Nat. Neurosci.* 6, 774–781. doi: 10.1038/nn1077
- Ohta, S., Fukui, N., and Sakai, K. L. (2013a). Computational principles of syntax in the regions specialized for language: integrating theoretical linguistics and functional neuroimaging. *Front. Behav. Neurosci.* 7:204. doi: 10.3389/fnbeh.2013.00204
- Ohta, S., Fukui, N., and Sakai, K. L. (2013b). Syntactic computation in the human brain: the Degree of Merger as a key factor. *PLoS ONE* 8:e56230. doi: 10.1371/journal.pone.0056230
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Poeppl, D., Emmorey, K., Hickok, G., and Pyllkkänen, L. (2012). Towards a new neurobiology of language. *J. Neurosci.* 32, 14125–14131. doi: 10.1523/jneurosci.3244-12.2012
- Radford, A. (2009). *Analysing English Sentences: A Minimalist Approach*. Cambridge: Cambridge University Press.
- Saito, M., and Fukui, N. (1998). Order in phrase structure and movement. *Linguist. Inq.* 29, 439–474. doi: 10.1162/002438998553815
- Sakai, K. L. (2005). Language acquisition and brain development. *Science* 310, 815–819. doi: 10.1126/science.1113530
- Sekerina, I. A. (2003). "Scrambling and processing: dependencies, complexity, and constraints," in *Word Order and Scrambling*, ed. S. Karimi (Malden, MA: Blackwell Publishing), 301–324.
- Stromswold, K., Caplan, D., Alpert, N., and Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain Lang.* 52, 452–473. doi: 10.1006/brln.1996.0024
- Suzuki, K., and Sakai, K. L. (2003). An event-related fMRI study of explicit syntactic processing of normal/anomalous sentences in contrast to implicit syntactic processing. *Cereb. Cortex* 13, 517–526. doi: 10.1093/cercor/13.5.517
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- Yasunaga, D., Yano, M., Yasugi, Y., and Koizumi, M. (2015). Is the subject-before-object preference universal? An event-related potential study in the Kaqchikel Mayan language. *Lang. Cogn. Neurosci.* 30, 1209–1229. doi: 10.1080/23273798.2015.1080372

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Ohta, Koizumi and Sakai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Grammatical Role Parallelism Influences Ambiguous Pronoun Resolution in German

Antje Saueremann\* and Natalia Gagarina\*

Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany

## OPEN ACCESS

### Edited by:

Aritz Irurtzun,  
Centre National de la Recherche  
Scientifique (CNRS), France

### Reviewed by:

Pei-Shu Tsai,  
National Changhua University  
of Education, Taiwan  
Maia Duguine,  
University of the Basque Country,  
Spain

### \*Correspondence:

Antje Saueremann  
saueremann@leibniz-zas.de  
Natalia Gagarina  
gagarina@leibniz-zas.de

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

Received: 02 December 2016

Accepted: 03 July 2017

Published: 25 July 2017

### Citation:

Saueremann A and Gagarina N (2017)  
Grammatical Role Parallelism  
Influences Ambiguous Pronoun  
Resolution in German.  
Front. Psychol. 8:1205.  
doi: 10.3389/fpsyg.2017.01205

Previous research on pronoun resolution in German revealed that personal pronouns in German tend to refer to the subject or topic antecedents, however, these results are based on studies involving subject personal pronouns. We report a visual world eye-tracking study that investigated the impact of the word order and grammatical role parallelism on the online comprehension of pronouns in German-speaking adults. Word order of the antecedents and parallelism by the grammatical role of the anaphor was modified in the study. The results show that parallelism of the grammatical role had an early and strong effect on the processing of the pronoun, with subject anaphors being resolved to subject antecedents and object anaphors to object antecedents, regardless of the word order (information status) of the antecedents. Our results demonstrate that personal pronouns may not in general be associated with the subject or topic of a sentence but that their resolution is modulated by additional factors such as the grammatical role. Further studies are required to investigate whether parallelism also affects offline antecedent choices.

**Keywords:** pronoun resolution, parallelism, grammatical role, word order, German

## INTRODUCTION

Pronoun resolution has “traditionally” been examined separately by linguists and psychologists. Yet, more recently both areas have come closer together. This led to the insight that anaphor/pronoun resolution is influenced by several factors. More importantly, the eye-tracking technique in the visual world paradigm has been shown to be particularly useful to examine pronoun/anaphor resolution during online processing. In this paradigm, an auditory stimulus is presented together with visual stimuli (e.g., two pictures) with the eye-movements on the pictures reflecting pronoun resolution preferences. Crucially, the online technique may reveal factors that influence pronouns resolution during online processing that may not be detected when offline techniques, e.g., judgments, are used (Schumacher et al., 2017). We used the visual world paradigm to investigate the impact of grammatical role parallelism which may likely to occur during online processing, i.e., exactly when the pronoun is processed (Smyth, 1994).

## Factors Influencing Pronoun Resolution

The factors influencing pronoun resolution have been intensively investigated in the last few decades. Pronoun resolution usually involves a process wherein an anaphor [e.g., the pronoun “he” in (1)] is associated with an antecedent in the previous context (e.g., “Goofy”).

(1) Goofy greets Donald. He...

Pronoun resolution requires the integration of different sources of information (e.g., Smyth, 1994; Arnold et al., 2000; Kehler et al., 2008; Schumacher et al., 2015, 2016, 2017).

First, syntactic factors, e.g., gender and number agreement and binding principles, constrain pronoun resolution. Second, different strategies may influence pronoun resolution in ambiguous contexts like (1), where the personal pronoun *he* may refer to *Goofy* or *Donald*, but participants usually prefer *Goofy*.

Resolution preferences in ambiguous contexts are influenced by the information status of the antecedent, i.e., personal pronouns refer to the most salient referents (e.g., Gundel et al., 1993; Ariel, 2001). The salience of an antecedent may be induced by several factors, among them its grammatical role (e.g., Crawley et al., 1990; Stevenson et al., 1994; Grosz et al., 1995; Bosch et al., 2007), thematic role (e.g., Schumacher et al., 2015, 2016, 2017), sentence position (Gernsbacher and Hargreaves, 1989; Stevenson et al., 1994) or information and discourse status (e.g., Grosz et al., 1995; Bosch et al., 2003; Bosch and Umbach, 2007; Kaiser and Trueswell, 2008; Colonna et al., 2012; Ellert, 2013).

The impact of one or the other factor from this list is usually difficult to disentangle. In addition, these factors interact with parallelism (e.g., Smyth, 1994; Stevenson et al., 1995; Chambers and Smyth, 1998), verb semantics (e.g., Grober et al., 1978; Schumacher et al., 2015, 2016, 2017), discourse relations (e.g., Grober et al., 1978; Kehler et al., 2008; see also Kaiser, 2011) and the type of referring expression realizing the anaphor (e.g., Gundel et al., 1993; Ariel, 2001; Bosch et al., 2003; Bosch et al., 2007; Kaiser and Trueswell, 2008; Schumacher et al., 2015).

Accordingly, different factors may be responsible for the subject preference of the anaphoric pronouns in the subject position, like, e.g., *he* in (1). These factors are difficult to tease apart because their features overlap and they make similar predictions. That is, *Goofy* may be the preferred antecedent because it is the subject and topic or because it shares the grammatical role and initial sentence position of the anaphor *he*. Languages with a more flexible word order than English, for instance German, provide a means to disentangle these factors. The goal of our study is to examine the impact of the word order of the antecedent sentence and grammatical role parallelism on pronoun resolution in German by using the visual world paradigm. First, we will review the research on pronoun resolution in German and then we will present our study. A discussion and conclusion will close the paper.

## Pronoun Resolution in German

German is a language with a relatively flexible word order, that allows besides the canonical SVO word order (2a) also the non-canonical OVS word order (2b). Word order variation has been linked to the information structure factors, in that the sentence-initial position is usually seen as a topic position (e.g., Frey, 2006). Thus, in SVO sentences the subject is seen as the topic and in OVS sentences the object.

(2) a) Der Mann grüßt den Jungen. Er/Der...  
 the<sub>NOM</sub><sup>1</sup> man greets the<sub>ACC</sub> boy he<sub>PRO</sub>/he<sub>DEM</sub>

<sup>1</sup>Note that in German determiners are also marked with respect to gender, but we did not indicate the gender in the glosses. In this example as well as in the items of our study, both the antecedents and the anaphora are masculine.

(2) b) Den Mann grüßt der Junge. Er/Der...  
 the<sub>ACC</sub> man greets the<sub>NOM</sub> boy he<sub>PRO</sub>/he<sub>DEM</sub>

Studies on pronoun resolution in German have mainly dealt with personal pronouns like *er* (“he<sub>PRO</sub>”) and demonstrative pronouns (d-pronouns) like *der* (“he<sub>DEM</sub>”) in contexts with the SVO (2a) or OVS (2b) word order of the antecedents.

Research on SVO sentences has shown that *er* (“he<sub>PRO</sub>”) is usually resolved to *der Mann* (“the man”) and *der* (“he<sub>DEM</sub>”) to *den Jungen* (“the boy”) (e.g., Bosch et al., 2003; Bouma and Hopp, 2007; Colonna et al., 2012; Schumacher et al., 2015, 2016, 2017; but see Bosch et al., 2007), but with a stronger preference for d-pronouns compared to pronouns (e.g., Bosch et al., 2003, 2007; Bouma and Hopp, 2007; Schumacher et al., 2015, 2017).

For OVS sentences the pattern is less coherent. While d-pronouns show a preference to refer to objects (Schumacher et al., 2015, 2017), personal pronouns prefer subjects (e.g., Bouma and Hopp, 2007; Schumacher et al., 2017) or show no preference (Bosch and Umbach, 2007; Bosch et al., 2007; Colonna et al., 2012; Schumacher et al., 2015). This variation in the pronoun resolution may be due to differences in the experimental material and settings, i.e., in the use of verbs (e.g., Schumacher et al., 2015, 2016, 2017), in the discourse relations between both sentences (e.g., Kaiser, 2011) or in the presence (or absence) of a preceding context which licenses the non-canonical word order (cf., Schumacher et al., 2017). In addition, differences in the methods used, especially in the use of offline or online experiments, may have led to incoherent results (cf., Schumacher et al., 2017; see also Bosch et al., 2007).

Despite this variation in the results, the majority of studies agree that subject personal pronouns usually refer to the subject (e.g., Bosch et al., 2007) or topic antecedent and indicate topic continuity (Bosch et al., 2003; Bosch and Umbach, 2007; Schumacher et al., 2016) whereas d-pronouns refer to non-subjects (e.g., Bosch et al., 2007) or less-topical referents and indicate a topic shift (Bosch et al., 2003; Bosch and Umbach, 2007; Schumacher et al., 2016). In addition, thematic status, subject status and information status have separate effects on pronoun resolution, which can be revealed when different constructions are investigated (e.g., Ellert, 2013; Schumacher et al., 2015, 2016).

While the resolution of subject pronouns has been explored much more intensively, only a few studies have investigated the resolution of pronouns with other grammatical roles, e.g., object pronouns. In these cases (e.g., Crawley et al., 1990; Smyth, 1994; Stevenson et al., 1994, 1995; Chambers and Smyth, 1998; Wolf et al., 2004; Kehler et al., 2008), researchers mainly examined English and used parallel structures like those in (3).

(3) Goofy greets Donald, and Daisy hugs him.

For these structures, some studies showed that the object pronoun *him* was associated with the object antecedent *Donald* (e.g., Smyth, 1994; Stevenson et al., 1995; Chambers and Smyth, 1998; Wolf et al., 2004; Kehler et al., 2008). However, other studies failed to provide evidence for this preference (e.g., Crawley et al., 1990; Stevenson et al., 1994), indicating that the resolution is

influenced by additional factors like verb semantics (e.g., Grober et al., 1978) or discourse relations (e.g., Kehler et al., 2008).

Crucially, the previous studies on English examined structures with three types of parallelism: first, *grammatical role parallelism*, with respect to the grammatical role (i.e., *him* and *Donald* are both the object), a second, *positional parallelism*, with respect to the position (*him* and *Donald* both occur in the sentence-final position), and a third, *structural parallelism*, with respect to the similar structures of the antecedent and anaphora sentences. Especially, the third type has been shown to have a strong impact on sentence processing (e.g., Sheldon, 1974; Frazier et al., 1984; Carlson, 2001; Callahan et al., 2010; Poirier et al., 2012 on English; Weskott, 2003; Knoeferle and Crocker, 2009 on German).

With respect to anaphor resolution, Smyth (1994) and Stevenson et al. (1995) tried to disentangle these factors. Stevenson et al. (1995) provide evidence against a “pure” position effect in the resolution of subject pronouns. Smyth (1994) showed that parallelism of the grammatical role had a strong impact on pronoun resolution, but structural parallelism also had an effect. However, neither study tested structures in the non-canonical word order, which may provide a clearer way to untangle the position effect and parallelism with respect to the grammatical role.

### The Present Study

We report on a visual world eye-tracking study that aimed to examine the impact of the word order and grammatical role parallelism on the online comprehension of personal pronouns. In the visual world paradigm the linguistic material [see (4)] is presented together with pictures of the possible antecedents (Figure 1), with the looks to the pictures of the antecedents reflecting pronoun resolution preferences during online processing.

We presented the antecedents in the canonical SVO (4a, 4b) or the non-canonical (4c, 4d) word order, with the case morphology of the determiners of the noun phrases (NPs) indicating grammatical role and word order. Grammatical role parallelism effects were tested by presenting the anaphoric pronoun either as the subject (4a, 4c) or as the object (4b, 4d).

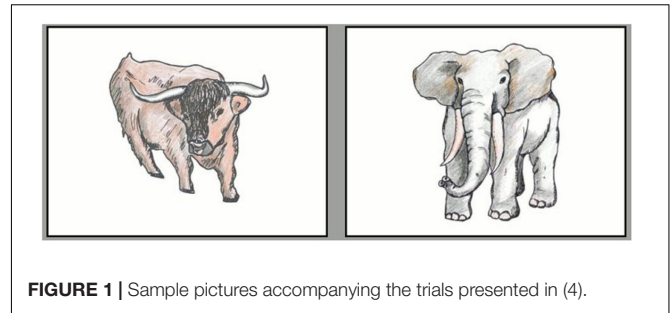


FIGURE 1 | Sample pictures accompanying the trials presented in (4).

This design, i.e., the comparison of subject and object anaphors, allows us to test the prediction that personal pronouns in general refer to the subject in the preceding antecedent sentence. If this is the case, we expect a higher proportion of looks to the picture of the subject antecedent compared to the object antecedent for both subject and object anaphora in the canonical word order (4a, 4b). With respect to the non-canonical word order (4c, 4d), the eye-tracking study by Schumacher et al. (2017) found a subject preference for subject anaphora (regardless of the word order) for accusative verbs, whereas offline studies revealed a less coherent pattern (e.g., Bosch et al., 2007; Schumacher et al., 2015). Given that our study is also an online study, we expect a subject preference for subject anaphora in our data.

If parallelism of a grammatical role plays a strong role during online processing, subject pronouns should be resolved to subject antecedents and object pronouns to object antecedents regardless of the word order of the antecedents. However, additional factors, e.g., positional and structural parallelism, may also play a role.

That is, if positional parallelism influences pronoun resolution, both subject and object pronouns should be resolved to the first mentioned antecedent in our study, i.e., to the subject in SVO sentences and the object in OVS sentences. That is, we expect an interaction between Pronoun Type and Word Order on the looks to the subject antecedents.

If structural parallelism influences pronoun resolution, we expect that subject pronouns are resolved to subject antecedents in SVO sentences (condition a) and object pronouns to object

(4) Der Bulle und der Elefant spielen zusammen Verstecken im Wald.  
 “The bull and the elephant are playing hide and seek in the forest.”

- |    |  |        |       |                    |            |                   |        |                    |           |            |
|----|--|--------|-------|--------------------|------------|-------------------|--------|--------------------|-----------|------------|
| a) | Der  | Bulle  | sieht | den                | Elefanten. | Er . . .          | ist    | traurig.           |           | (SVO, sbj) |
|    | the <sub>NOM</sub>   | bull   | sees  | the <sub>ACC</sub> | elephant   | he <sub>NOM</sub> | is     | sad                |           |            |
|    | “The bull sees the elephant. He . . . is sad.”               |        |       |                    |            |                   |        |                    |           |            |
| b) | Der  | Bulle  | sieht | den                | Elefanten. | Ihn . . .         | trifft | der                | Blitz.    | (SVO, obj) |
|    | the <sub>NOM</sub>   | bull   | sees  | the <sub>ACC</sub> | elephant   | he <sub>ACC</sub> | hits   | the <sub>NOM</sub> | lightning |            |
|    | “The bull sees the elephant. Him . . . the lightning hits.”  |        |       |                    |            |                   |        |                    |           |            |
| c) | Den  | Bullen | sieht | der                | Elefant.   | Er . . .          | ist    | traurig.           |           | (OVS, sbj) |
|    | the <sub>ACC</sub>   | bull   | sees  | the <sub>NOM</sub> | elephant   | he <sub>NOM</sub> | is     | sad                |           |            |
|    | “The bull, the elephant sees. He . . . is sad.”              |        |       |                    |            |                   |        |                    |           |            |
| d) | Den  | Bullen | sieht | der                | Elefant.   | Ihn . . .         | trifft | der                | Blitz.    | (OVS, obj) |
|    | the <sub>ACC</sub>   | bull   | sees  | the <sub>NOM</sub> | elephant   | he <sub>ACC</sub> | hits   | the <sub>NOM</sub> | lightning |            |
|    | “The bull, the elephant sees. Him . . . the lightning hits.” |        |       |                    |            |                   |        |                    |           |            |

antecedents in OVS sentences (condition d). In the conditions without structural parallelism (conditions 4b and 4c), we expect less clear resolution preferences.

## MATERIALS AND METHODS

### Design and Materials

The experiment employed a  $2 \times 2$  repeated-measures design with Word Order (SVO vs. OVS) and (the grammatical role of the) Pronoun (“subject” (sbj) vs. “object” (obj)) as independent variables and the eye-movements, i.e., the proportion of looks to the subject of the SVO or OVS sentence, as dependent variable.

The experimental trials [see (4)] started with a sentence introducing the two referents, which was followed by an antecedent SVO or OVS word order sentence. The grammatical role of the antecedents was indicated by case marking of the first and second NP: the determiner *der* indicated nominative case and subject status, and the determiner *den* indicated accusative case and object status. The antecedent sentence was followed by a second sentence with the subject pronoun *er* (“he”) or object pronoun *ihn* (“him”) in the initial position. The pronoun sentence was interrupted by a pause of 500 ms after the offset of the pronoun.

The verbal stimulus was accompanied by two pictures depicting the two animals mentioned in the discourse (see **Figure 1** above). The pictures had a size of 440 pixels  $\times$  330 pixels and were placed horizontally at the left or right side of the screen, separated from each other by approximately 25 pixels.

Four experimental items (animal pairs) were created (see Supplementary Material for the complete list of the items). For each item two versions of the trials were created controlling for the effects of order of mention and positioning of the pictures. That is, for each trial we created an alternative version wherein the elephant was the first NP in the lead-in and antecedent sentence and the picture was presented on the left side. Each participant saw all four conditions of an item. The reason for this experimental design and the low number of items was that the experiment was also run with bilingual preschoolers, who should know the meanings of the verbs used.

In addition two practice trials and eight filler sentences were created. Each trial was accompanied by two pictures of two animals. Practice trials consisted of an introduction sentence and a transitive sentence, similar to the SVO condition of the experimental trials. However, these sentences were not followed by a pronoun sentence. Fillers were SVO sentences that mentioned the two animals depicted.

### Procedure

Participants were seated in front of a 15" laptop on which the experimental sentences were presented. The experiment involved a looking-while-listening task. That is, participants were not instructed to perform a specific task but only to listen to short stories that were accompanied by two pictures.

Each experimental session began with a 5-point calibration procedure to adjust the eye-tracking system. The experiment started with two practice sentences. Each participant saw 16

experimental trials, with a filler sentence being shown after every two experimental trials. Participants were tested using four test lists that were created to control for the positioning of the pictures and the order of the mention of the animals.

Data were recorded using a portable Tobii X2-60 Compact eye-tracking system (Tobii Technology AB, Sweden), which was attached to the laptop. Eye-movements were sampled with a tracking rate of 60 Hz, approximately every 16 ms.

### Data Treatment and Analyses

The eye-movement recordings were based on the gazes as determined and pre-processed by the Tobii Studio software (Version 3.2.2, Tobii Technology AB, Sweden). Trials with more than 50 percent track loss (looks off screen) were excluded from further analysis (1%).

The eye-movement data was aggregated in 50 ms bins and analyzed in twelve 250 ms time windows from the onset of the pronoun until the end of the sentence. For the statistical analyses, we calculated the empirical logit for the looks to the picture of the subject antecedent, aggregating over items (cf., Barr, 2008). Looks to the subject antecedent picture were almost complementary to looks to the object antecedent picture because looks to neither of the pictures were rare (2%).

The lme4 package (version 1.1-12; Bates et al., 2015) was used to calculate linear mixed-effects models to assess the fixed effects of Word Order, Pronoun, Time and their interactions, and the random effect of Participants on the empirical logit of the looks to the target picture. The models included the weightings recommended for empirical logit analyses (Barr, 2008). The specification of the random effects of Participants considered the slope adjustment for Pronoun and Word Order and their interaction (cf., Barr et al., 2013). Time was not considered for the slope adjustment because models that included Time for slope adjustment led to convergence errors.

The contrast codings of predictors and Word Order (SVO: +1, OVS: -1) and Pronoun (er: +1, ihn: -1) and their interaction resembled those of traditional ANOVA analyses. The continuous predictor Time captured the five time (50 ms) bins that were analyzed in each 250 ms time window.

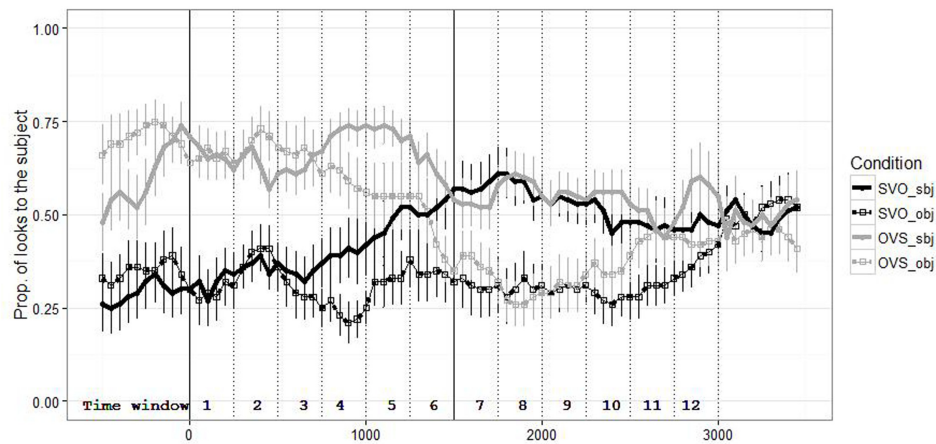
### Participants

Eighteen students of the Humboldt University Berlin participated in the study (13 women, mean age: 27 years). They were monolingual native speakers of German and had normal or corrected to normal vision.

This study was carried out in accordance with the recommendations of the Declaration of Helsinki with written informed consent from all subjects. All subjects gave written informed consent prior to participation in accordance with the Declaration of Helsinki. The protocol was approved by the German Linguistic Society (Deutsche Gesellschaft für Sprachwissenschaft, DGfS)

## RESULTS

**Figure 2** shows the mean proportion of looks to the subject calculated on 50 ms time bins starting with the offset of the



**FIGURE 2 |** Mean proportion (with SE) of looks to the subject antecedent depending on Word Order and Pronoun. Standard errors (SE) exclude between-participant variance (Cousineau, 2005) and were normalized using Morey's (2008) correction. Note that looks to the object antecedent were complementary to the looks to the subject.

antecedent sentence (SVO vs. OVS). The proportions of looks following SVO sentences are shown in black color and those following OVS sentences in gray. Solid lines indicate trials with the subject pronoun and dotted lines those with the object pronoun. The solid vertical lines indicate the onset of the pronoun (*er* or *ihn*) and the onset of the continuation of the sentence. Dotted vertical lines indicate the time windows.

**Table 1** lists the intercept ( $b$ ) and  $t$ -values ( $t$ ) for the fixed effects of the models in each time window. The models revealed a significant effect of word order in the first five time windows (until 1250 ms), resulting from fewer looks to the subject following SVO sentences whereas there were more looks to the subject following OVS sentences. The effect gradually declined in the fifth and sixth time windows, as indicated by the Pronoun–Time interaction, and did not occur in the subsequent time windows. We propose that this eye-movement pattern reflects the looks to the last-mentioned referent of the transitive sentence, i.e., the subject in OVS and the object in SVO sentences.

The pronoun type influenced the eye-movement from around 750 ms (starting with the fourth time window), as a significant interaction between Pronoun and Time revealed.<sup>2</sup> This interaction indicates that the difference between subject and object anaphora increased with time. That is, looks to the subject antecedent gradually increased after subject anaphora (solid gray and black lines) and gradually decreased after object anaphora (dotted gray and black lines) during the time interval from 750–1000 ms, i.e., in the fourth time window. Notably this effect occurred in both word orders. The main effect of Pronoun was fully established in the fifth time window (from 1000 ms) and continued until the tenth time window (until 2500 ms). In the eleventh time window (2500–2750 ms), the Pronoun effect

<sup>2</sup>The models revealed an interaction between Time and Pronoun type in the second time window between 250 and 500 ms. Note, however, that it takes around 200 ms to initiate a saccade (e.g., Sumner, 2011) and around 400 ms to utter the pronouns. Accordingly, we do not expect the pronoun to have already had an effect in this time window.

gradually disappeared, as indicated by the interaction between Pronoun and Time. In the final time window (2750–3000 ms), there was a significant interaction between Pronoun, Word Order and Time as well as a main effect of Time. *Post hoc* comparisons assessing the impact of Time and Word Order for each pronoun type revealed a significant effect of Time reflecting a gradual increase in the eye-movements for subject anaphora ( $b = 0.002$ ,  $t = 3.431$ , especially in OVS trials) but no change in the eye-movements for object anaphora ( $b = 0.000$ ,  $t = 0.691$ ). However, given that this time window was at the end of the trial and the eye-movements in all conditions centered around chance-level, the effects in the last two time windows are difficult to interpret.

## DISCUSSION

The eye-tracking study examined the effects of word order and grammatical role parallelism on anaphora resolution in adult German. Antecedent sentences with SVO and OVS word order and sentences with subject vs. object pronominal anaphora composed four contexts which were investigated [see examples in (4)].

The results showed that grammatical role parallelism influenced online pronoun resolution in both word orders. This was reflected by the eye-movements starting around 750 ms after pronoun onset such that looks to the subject antecedent increased in subject anaphora trials compared to object anaphora trials in both word orders. Given that looks to subject and object antecedents were complementary, this also reflects that looks to the object increased after object anaphora trials compared to subject anaphora trials. This pattern occurred even before the anaphora sentence was continued, suggesting that it cannot be attributed to the different sentence continuations for subject and object anaphora.

Importantly, this effect of the pronoun type was not influenced by an interaction with word order. This suggest

**TABLE 1 |** Fixed effects of the models predicting the looks to the subject picture (significant values at  $\alpha = 0.05$ ,  $|t| \geq 2$  are indicated in bold).

	1 (0–250)		2 (250–500)		3 (500–750)		4 (750–1000)		5 (1000–1250)		6 (1250–1500)	
	<i>b</i>	<i>t</i>	<i>b</i>	<i>t</i>	<i>b</i>	<i>t</i>	<i>b</i>	<i>t</i>	<i>b</i>	<i>t</i>	<i>b</i>	<i>t</i>
Intercept	−0.104	−0.681	−0.055	−0.294	0.008	0.049	−0.066	−0.468	−0.084	−0.544	0.076	0.470
Pronoun	0.081	0.679	0.012	0.117	−0.045	−0.333	0.227	1.643	<b>0.430</b>	<b>2.789</b>	<b>0.397</b>	<b>2.151</b>
Word order	<b>−0.887</b>	<b>−4.956</b>	<b>−0.706</b>	<b>−3.809</b>	<b>−0.690</b>	<b>−3.600</b>	<b>−0.811</b>	<b>−4.212</b>	<b>−0.714</b>	<b>−3.126</b>	−0.367	−1.998
Time	0.000	0.283	<b>0.001</b>	<b>2.416</b>	−0.000	−1.015	0.000	0.179	<b>0.001</b>	<b>2.195</b>	<b>−0.001</b>	<b>−4.043</b>
Pron × WO*	−0.046	−0.302	0.011	0.130	0.126	0.996	0.057	0.459	0.009	0.054	0.048	0.280
Pron × Time	0.000	0.520	<b>−0.001</b>	<b>−2.808</b>	0.001	1.870	<b>0.001</b>	<b>2.863</b>	0.000	0.083	0.000	1.283
WO × Time	0.001	1.645	0.000	1.036	<b>−0.001</b>	<b>−2.047</b>	0.000	0.423	<b>0.001</b>	<b>3.522</b>	<b>0.001</b>	<b>3.234</b>
Pron × WO × Time	0.000	0.789	0.000	0.647	−0.000	−0.209	−0.000	−0.632	0.000	0.762	0.000	0.335
	7 (1500–1750)		8 (1750–2000)		9 (2000–2250)		10 (2250–2500)		11 (2500–2750)		12 (2750–3000)	
	<i>b</i>	<i>t</i>	<i>b</i>	<i>t</i>	<i>b</i>	<i>t</i>	<i>b</i>	<i>t</i>	<i>b</i>	<i>t</i>	<i>b</i>	<i>t</i>
Intercept	−0.271	−1.347	−0.177	−1.574	−0.346	−2.929	−0.315	−2.387	−0.340	−2.384	−0.291	−2.053
Pronoun	<b>0.582</b>	<b>2.724</b>	<b>0.715</b>	<b>4.311</b>	<b>0.570</b>	<b>3.957</b>	<b>0.470</b>	<b>2.929</b>	0.296	1.567	0.186	1.159
Word order	−0.086	−0.500	0.038	0.264	−0.001	−0.006	−0.029	−0.221	−0.230	−1.611	−0.133	−0.899
Time	−0.000	−1.226	−0.001	−1.538	0.000	1.057	−0.000	−1.397	−0.000	−0.124	<b>0.001</b>	<b>2.844</b>
Pron × WO	0.128	0.926	0.059	0.463	−0.032	−0.307	0.017	0.171	0.041	0.314	0.106	0.779
Pron × Time	0.000	0.732	−0.000	−1.221	−0.000	−0.901	−0.000	−0.224	<b>−0.001</b>	<b>−2.486</b>	0.001	1.780
WO × Time	0.000	0.661	0.000	0.270	−0.000	−1.124	<b>−0.001</b>	<b>−2.788</b>	0.000	0.878	0.000	0.246
Pron × WO × Time	0.000	0.453	−0.001	−1.665	0.000	0.471	−0.000	−0.703	0.001	1.710	<b>−0.001</b>	<b>−3.038</b>

\*Pron, pronoun; WO, word order.

that the resolution preferences resulted from parallelism of the grammatical role and were not restricted to a particular position of an antecedent or to similarities of the syntactic structure of the antecedent and anaphor sentence. Thus, the pronoun resolution in our study was not influenced by positional or structural parallelism.

Nevertheless, the eye-movements were initially also influenced by the word order of the sentence, reflecting that participants looked at the last-mentioned antecedent. This effect did not interact with the grammatical role of the anaphora and apparently resulted from the experimental design. In addition, in later time windows when the sentence continued, the word order effect gradually decreased and did not affect eye-movements.

Our results strongly indicate that the grammatical role of the anaphor influences its resolution shortly after the pronoun is heard and processed and even before the anaphor sentence is continued. Indeed, the time window wherein the impact of the anaphora occurred in our study corroborates the results of the visual world study by Schumacher et al. (2017), who found an impact of the demonstrative and personal pronouns on their resolution in accusative verb sentences only slightly earlier (400–600 ms after pronoun onset). This suggests that not only the type of the referring expression but also the grammatical role impacts online pronoun resolution.

The early effect of the grammatical role of the pronoun corresponds to the proposal by Smyth (1994). Similar to previous research concluding that pronoun resolution starts immediately after the pronoun is heard (e.g., Ehrlich and Rayner, 1983; see also Arnold et al., 2000; Schumacher et al., 2016, 2017), Smyth suggests that parallelism influences pronoun resolution

in terms of a feature match process whereby antecedents are selected on the basis of the features they share with the anaphora – in our case, grammatical role features. In our study, this effect was not restricted to structures in which the antecedent sentence and the anaphor sentence share the same word order, i.e., positional parallelism. This differs from the studies demonstrating a strong impact of positional (or structural) parallelism on sentence processing (e.g., Sheldon, 1974; Frazier et al., 1984; Carlson, 2001; Weskott, 2003; Knoeferle and Crocker, 2009; Callahan et al., 2010), including pronoun resolution (e.g., Smyth, 1994; Poirier et al., 2012). This difference may result from the materials (e.g., the lack of a conjunction or the pause within the pronoun sentence in our study) or the methodology used.

While our data also show a stable effect of the pronoun, reflecting the grammatical role parallelism effect, until 2250 ms after the pronoun onset, this did not influence eye-movements in the last two time regions. The lack of the effect in these time regions may merely result from the fact that they appear at the very end of the trial. Alternatively, it may indicate that grammatical role parallelism effects may be weaker during later processing or influenced by the predicate of an anaphora sentence. This instability with respect to the resolution preferences was also found in Schumacher et al.’s (2017) research and was evidenced by the differences between their online and offline study. While their online eye-tracking study revealed a subject preference for subject personal pronouns in SVO and OVS sentences (Schumacher et al., 2017), their offline rating study showed the subject preference in SVO sentences only (Schumacher et al., 2015).



Given that we did not test offline antecedent choice, we can only draw cautious predictions about the offline interpretation of the subject and object personal pronouns in our data. Nevertheless, our results reflect a stable effect of the grammatical role. This suggests that personal pronouns in the initial position in a sentence are not generally – irrespective of the other factors – resolved to subjects but that their resolution preferences are also modulated by the grammatical role parallelism of a pronominal anaphora and its antecedent. This corresponds to previous work (e.g., Bosch and Umbach, 2007; Bosch et al., 2007; Schumacher et al., 2015, 2017) demonstrating that (subject) personal pronouns show weaker antecedent preferences compared to demonstrative pronouns.

Notably, visual inspection of the eye-movement plot may indicate that the impact of the grammatical role was somewhat stronger for object anaphora compared to subject anaphora because the eye-movements for subject anaphora were closer to the 50% chance level. This apparently weak preference for subject anaphora also corresponds to the differences between personal and demonstrative pronouns mentioned above. However, this does not explain why object anaphora show a clearer preference for object antecedents.

It might be that hearers rely more on parallelism when the object pronoun follows the less frequent and more marked OVS word order in the antecedent sentence. Following the SVO sentence, the OVS sentences with an object anaphor may indicate a topic shift with the object as the new topic. Following the OVS antecedent sentence, structural parallelism with the OVS sentence may facilitate OVS sentence comprehension in general (cf., Weskott, 2003; Knoeferle and Crocker, 2009) and thus may enhance grammatical role parallelism effects. If this is the case, parallelism effects may interact with information structure factors. However, further research that considers corpus data and antecedent choice tasks is needed to clarify the differences between subject and object anaphora.

In general, our study underlines the importance of considering different empirical methods in the study of pronoun interpretation. We employed the eye-tracking method within the visual world paradigm wherein the eye-gazes to the pictures reflect pronoun resolution during online processing. Yet, this method does not only provide insight into the different sources of information considered during online comprehension but is also an implicit measure of sentence comprehension which reduces task demands especially for children (e.g., Brandt-Kobele and Höhle, 2010; Bergmann et al., 2012). However, the technique also has its limitations. The online results may not always correspond to offline responses (Schumacher et al., 2017) because they do not capture processes during later stages of sentence processing/interpretation. Furthermore, the method may be more time-consuming compared to offline methods regarding to the creation of the experimental materials (visual and auditory material) and the preprocessing and the analyses of eye-movements.

In addition, our study underlines that research on pronoun resolution (or more general language use/production and comprehension) should consider both linguistic and psycholinguistic approaches. In particular, our study demonstrates that,

in addition to the linguistic factors (e.g., agreement, personal pronoun vs. d-pronoun), processing factors like grammatical role parallelism influence pronoun resolution. In this way, it emphasizes the requirement that linguistic theories should be based on empirical work that employs different methods.

## CONCLUSION

To summarize, we reported on the first study comparing the impact of word order and parallelism effects on online pronoun resolution in German. We showed that parallelism of the grammatical role had an early and strong effect on the processing of the pronoun, regardless of the word order of the antecedents. This suggests that different sources of information are considered during online pronoun resolution (cf., Arnold et al., 2000; Kehler et al., 2008; Schumacher et al., 2016, 2017) and that parallelism is one of the crucial factors in this process (cf., Smyth, 1994). In addition, our results indicate that personal pronouns may not in general be associated with the subject or topic of a sentence in German but that their resolution is modulated by additional factors such as the grammatical role. Further studies are required to investigate whether parallelism also affects offline antecedent choices and whether the parallelism may also influence pronoun resolution of demonstrative pronouns. In this way, the interaction between parallelism and information structure may be clarified.

## AUTHOR CONTRIBUTIONS

NG was responsible for the design and procedure of the experiments. NG supervised a Ph.D.-student, Elena Valentik-Klein, who is not working in research anymore, during the creation the materials. The experiment was carried out by student researchers. AS conducted the data analyses. AS and NG co-authored the paper.

## FUNDING

This work was supported (in part) by Bundesministerium für Bildung und Forschung (BMBF) (Grant Nr. 01UG1411). The publication of this article was funded by the Open Access Fund of the Leibniz Association.

## ACKNOWLEDGMENTS

We would like to thank to Elena Valentik-Klein, Anna Konik, Anne la Porte, Levke Schneekloth and Natalie Suerlemi for help in designing and creating the experiments as well as in the data collection. We also thank Ilka Lörke for painting the pictures.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01205/full#supplementary-material>

## REFERENCES

- Ariel, M. (2001). "Accessibility theory: an overview," in *Text Representation: Linguistic and Psycholinguistic Aspects*, eds T. Sanders, J. Schliperoord, and W. Spooren (Amsterdam: John Benjamins), 29–87. doi: 10.1075/hcp.8.04ari
- Arnold, J. E., Eisenband, J. E., Brown-Schmidt, S., and Trueswell, J. C. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition* 76, B13–B26. doi: 10.1016/S0010-0277(00)00073-1
- Barr, D. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *J. Mem. Lang.* 59, 457–474. doi: 10.1016/j.jml.2007.09.002
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bergmann, C., Paulus, M., and Fikkert, P. (2012). Preschoolers' comprehension of pronouns and reflexives: the impact of the task. *J. Child Lang.* 39, 777–803. doi: 10.1017/S0305000911000298
- Bosch, P., Katz, G., and Umbach, C. (2007). "The non-subject bias of German demonstrative pronouns," in *Anaphors in Text: Cognitive, Formal and Applied Approaches to Anaphoric Reference*, eds M. Schwarz-Friesel, M. Consten, and M. Knees (Amsterdam: John Benjamins), 145–164. doi: 10.1075/slcs.86.13bos
- Bosch, P., Rozario, T., and Zhao, Y. (2003). "Demonstrative pronouns and personal pronouns. German der vs. er," in *Proceedings of the EACL 2003: Workshop on the Computational Treatment of Anaphora*, Budapest.
- Bosch, P., and Umbach, C. (2007). Reference determination for demonstrative pronouns. *ZAS Pap. Linguist.* 48, 39–51.
- Bouma, B., and Hopp, H. (2007). Coreference preferences for personal pronouns in German. *ZAS Pap. Linguist.* 48, 53–74.
- Brandt-Kobe, O.-C., and Höhle, B. (2010). What asymmetries within comprehension reveal about asymmetries between comprehension and production: the case of verb inflection in language acquisition. *Lingua* 120, 1910–1925. doi: 10.1016/j.lingua.2010.02.008
- Callahan, S. M., Shapiro, L. P., and Love, T. (2010). Parallelism effects and verb activation: the sustained reactivation hypothesis. *J. Psycholinguist. Res.* 39, 101–118. doi: 10.1007/s10936-009-9128-0
- Carlson, K. (2001). The effects of parallelism and prosody in the processing of gapping structures. *Lang. Speech* 44, 1–26. doi: 10.1177/00238309010440010101
- Chambers, C. G., and Smyth, R. (1998). Structural parallelism and discourse coherence: a test of centering theory. *J. Mem. Lang.* 39, 593–608. doi: 10.1006/jmla.1998.2575
- Colonna, S., Schimke, S., and Hemforth, B. (2012). Information structure effects on anaphora resolution in German and French: a crosslinguistic study of pronoun resolution. *Linguistics* 50, 901–1013. doi: 10.1515/ling-2012-0031
- Cousineau, D. (2005). Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. *Tutor. Quant. Methods Psychol.* 1, 42–45. doi: 10.3758/s13423-012-0230-1
- Crawley, R. A., Stevenson, R. J., and Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *J. Psycholinguist. Res.* 19, 245–264. doi: 10.1007/BF01077259
- Ehrlich, K., and Rayner, K. (1983). Pronoun assignment and semantic integration during reading: eye movements and immediacy of processing. *J. Verbal Learn. Verbal Behav.* 22, 75–87. doi: 10.1016/S0022-5371(83)80007-3
- Ellert, M. (2013). Information structure affects the resolution of the subject pronoun er and der in spoken German discourse. *Discours* 12, 1–24. doi: 10.4000/discours.8756
- Frazier, L., Taft, L., Roeper, T., Clifton, C. E., and Ehrlich, K. (1984). Parallel structure: a source of facilitation in sentence comprehension. *Mem. Cogn.* 12, 421–430. doi: 10.3758/BF03198303
- Frey, W. (2006). "Contrast and movement to the German prefield," in *The Architecture of Focus*, eds V. Molnár and S. Winkler (Berlin: de Gruyter), 235–264.
- Gernsbacher, M. A., and Hargreaves, D. J. (1989). Accessing sentence participants: the advantage of first mention. *J. Mem. Lang.* 27, 699–717. doi: 10.1016/0749-596X(88)90016-2
- Grober, E. H., Beardsley, W., and Caramazza, A. (1978). Parallel function strategy in pronoun assignment. *Cognition* 6, 117–133. doi: 10.1016/0010-0277(78)90018-5
- Grosz, B. J., Joshi, A. K., and Weinstein, S. (1995). Centering: a framework for modeling the local coherence of discourse. *Comput. Linguist.* 21, 203–225.
- Gundel, J. K., Hedberg, N., and Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language* 69, 274–307. doi: 10.2307/416535
- Kaiser, E. (2011). "On the relation between coherence relations and anaphoric demonstratives in German," in *Proceedings of the Sinn & Bedeutung 15*, eds I. Reich, E. Horch, and D. Pauly (Saarbrücken: Universitätsverlag des Saarlandes), 337–351.
- Kaiser, E., and Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: evidence for a form-specific approach to reference resolution. *Lang. Cogn. Process.* 23, 709–748. doi: 10.1016/j.cognition.2009.03.010
- Kehler, A., Kertz, L., Rohde, H., and Elman, J. (2008). Coherence and coreference revisited. *J. Semant.* 25, 1–44. doi: 10.1093/jos/ffm018
- Knoeferle, P., and Crocker, M. W. (2009). Constituent order and semantic parallelism in online comprehension: eye-tracking evidence from German. *Q. J. Exp. Psychol.* 62, 2338–2371. doi: 10.1080/17470210902790070
- Morey, R. D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutor. Quant. Methods Psychol.* 4, 61–64. doi: 10.20982/tqmp.04.2.p061
- Poirier, J., Walenski, M., and Shapiro, L. P. (2012). The role of parallelism in the real-time processing of anaphora. *Lang. Cogn. Process.* 27, 868–886. doi: 10.1080/01690965.2011.601623
- Schumacher, P. B., Backhaus, J., and Dangl, M. (2016). Backward- and forward-looking potential of anaphors. *Front. Psychol.* 6:1746. doi: 10.3389/fpsyg.2015.01746
- Schumacher, P. B., Dangl, M., and Uzun, E. (2015). "Thematic role as prominence cue during pronoun resolution in German," in *Empirical Perspectives on Anaphora Resolution*, eds A. Holler and K. Suckow (Berlin: de Gruyter), 213–239.
- Schumacher, P. B., Roberts, L., and Järvi, J. (2017). Agentivity drives real-time pronoun resolution: evidence from German "er" and "der". *Lingua* 185, 25–41. doi: 10.1016/j.lingua.2016.07.004
- Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *J. Verbal Learn. Verbal Behav.* 13, 272–281. doi: 10.1016/S0022-5371(74)80064-2
- Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *J. Psycholinguist. Res.* 23, 197–229. doi: 10.1007/BF02139085
- Stevenson, R. J., Crawley, R. A., and Kleinman, D. (1994). Thematic roles, focus, and the representation of events. *Lang. Cogn. Process.* 9, 519–548. doi: 10.1186/2041-1480-3-12
- Stevenson, R. J., Nelson, A. W. R., and Stenning, K. (1995). The role of parallelism in strategies of pronoun comprehension. *Lang. Speech* 38, 393–418. doi: 10.1177/002383099503800404
- Sumner, P. (2011). "Determinants of saccadic latency," in *The Oxford Handbook of Eye Movements*, eds S. P. Liversedge, I. D. Gilchrist, and S. Everling (Oxford: Oxford University Press), 413–424.
- Weskott, T. (2003). *Information Structure as a Processing Guide: The Left Periphery of German Verb-Second Sentences and Its Interpretation in Context*. Doctoral thesis, University of Leipzig, Leipzig.
- Wolf, F., Gibson, E., and Desmet, T. (2004). Discourse coherence and pronoun resolution. *Lang. Cogn. Process.* 19, 665–675. doi: 10.1080/01690960444000034

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Sauermann and Gagarina. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Understanding Grammars through Diachronic Change

Nerea Madariaga \*

Faculty of Arts, University of the Basque Country (UPV/EHU), Vitoria-Gasteiz, Spain

In this paper, I will vindicate the importance of syntactic change for the study of synchronic stages of natural languages, according to the following outline. First, I will analyze the relationship between the diachrony and synchrony of grammars, introducing some basic concepts: the notions of I-language/E-language, the role of Chomsky's (2005) three factors in language change, and some assumptions about language acquisition. I will briefly describe the different approaches to syntactic change adopted in generative accounts, as well as their assumptions and implications (Lightfoot, 1999, 2006; van Gelderen, 2004; Biberauer et al., 2010; Roberts, 2012). Finally, I will illustrate the convenience of introducing the diachronic dimension into the study of at least certain synchronic phenomena with the help of a practical example: variation in object case marking of several verbs in Modern Russian, namely, the verbs denoting avoidance and the verbs *slušať'sja* "obey" and *dožidat'sja* "expect," which show two object case-marking patterns, genitive case in standard varieties and accusative case in colloquial varieties. To do so, I will review previous descriptive and/or functionalist accounts on this or equivalent phenomena (Jakobson, 1984 [1936]; Clancy, 2006; Nessel and Kuznetsova, 2015a,b). Then, I will present a formal—but just synchronic—account, applying Sigurðsson (2011) hypothesis on the expression of morphological case to this phenomenon. Finally, I will show that a formal account including the diachronic dimension is superior (i.e., more explanative) than purely synchronic accounts.

**Keywords:** syntactic change, Old Russian, Modern Russian, variation, object case marking, accusative case, genitive case

## OPEN ACCESS

### Edited by:

Aritz Irurtzun,  
Centre National de la Recherche  
Scientifique (CNRS), France

### Reviewed by:

Ricardo Etxepare,  
IKER UMR5478, Centre National de la  
Recherche Scientifique (CNRS),  
France

David W. Lightfoot,  
Georgetown University, United States

Anton V. Zimmerling,  
Moscow State Pedagogical University,  
Russia

### \*Correspondence:

Nerea Madariaga  
nerea.madariaga@ehu.es

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 03 May 2017

**Accepted:** 04 July 2017

**Published:** 31 July 2017

### Citation:

Madariaga N (2017) Understanding  
Grammars through Diachronic  
Change. *Front. Psychol.* 8:1226.  
doi: 10.3389/fpsyg.2017.01226

## INTRODUCTION

It seems a straightforward assumption to acknowledge diachronic change as the most important source of variation in languages and a crucial factor in shaping grammars. It is difficult not to agree with Lightfoot (in preparation) in that "nothing in syntax makes sense except in the light of change," paraphrasing, in turn, the famous adagio by Dobzhansky (1973) that "nothing in biology makes sense except in the light of evolution." Given the fact that most variable properties in languages emerge through change, it seems reasonable to include the relevant historical facts in any study on variation, at least in those cases when the history of the language concerned is sufficiently attested.

However, the role of historical linguistics does not receive the attention it deserves in synchronic studies. In this paper, I vindicate the importance of introducing the diachronic dimension into the formal study of at least certain synchronic phenomena, by highlighting the role of syntactic change through a specific example of variation in Russian. First, I analyze the relationship between diachrony and synchrony of grammars,

introducing some basic concepts: the notion of syntactic change, its abruptness and discreteness, the contrast between I-language and E-language, the relevance of language acquisition, and its role in syntactic change, as well as the effect of Chomsky's (2005) third factor in language change. Further, I describe the case alternation between genitive vs. accusative complements of certain medial verbs in present-day Russian (the so-called *-sja* verbs). Then, I review the shortcomings of purely synchronic accounts of different linguistic orientations applied to this specific case of variation. Finally, I prove that an account introducing the diachronic dimension can be explanatorily superior, at least, in this specific case study on variation. The final section contains some conclusions to this paper.

## BASIC NOTIONS ABOUT DIACHRONIC GENERATIVE SYNTAX

In this section, I will introduce some basic notions on historical change assumed by generative approaches to grammar (as opposed to other linguistic schools, mainly usage-based or functionalist approaches). Diachronic generative studies started in the early 70s, with Andersen's (1973) article on abductive change and Lightfoot's (1974) work on modals, preceded by Klima's (1965) dissertation (*Studies in diachronic transformational syntax*). The foundational work on diachronic generative syntax is unanimously considered to be Lightfoot's (1979) *Principles of diachronic syntax*; it gave rise to a productive research program within formal linguistic studies. As an example, see the collective volumes, the product of the biennial DIGS (*Diachronic Generative Syntax*) conference, published by OUP. Recently, CUP published the collective reference handbook on diachronic generative syntax *Cambridge handbook of historical syntax*, edited by Ledgeway and Roberts (2017).

A basic notion in generative approaches to diachrony is the view of syntactic change as a special kind of "reanalysis" or rather "new analysis," as firstly claimed by Lightfoot (1979 and subsequent work—1991, 1999, 2006), and later widely adopted in the generative linguistic community (Faarlund, 1990; Hale, 1998; Roberts and Roussou, 2003; van Gelderen, 2004; Roberts, 2007; etc.). Within this view, learners acquire a language by parsing or analyzing the relevant input, also called Primary Linguistic Data (PLD). Most of the time learners succeed in converging with the grammar/structure that generated its input, a property called *inertia* in formal grammar (Longobardi, 2001). Syntactic change, then, stems from a special type of "analysis" or "parsing" of the PLD a learner can perform during the language acquisition process; namely, in the case when, for some reason, the learner's grammar/structure does not converge with the grammar/structure that generated its input. This is known as the discontinuity or failure of transmission between generations.

In generative approaches to change, the discontinuity of transmission is usually assumed to be abrupt (rather than gradual), in the sense that grammars are acquired afresh by each speaker (Lightfoot, 1979, 1991 and afterward). What seems like gradual change is reduced in diachronic generative syntax to successive discrete changes according to the following

considerations: (i) A change can initially affect only specific items or structures, and then spread to more items or syntactic environments (van Gelderen, 2010; Madariaga, 2012). (ii) A change can spread through a linguistic community, giving rise to situations of diglossia and "competing grammars" (Kroch, 1989; Yang, 2002). (iii) A change can produce different synchronic variants coexisting in a single speaker at different linguistic levels, which we commonly call "I-language" vs. "E-language." I-language stands for the linguistic competence of each individual speaker, while E-language refers to the linguistic productions of a community of speakers (Chomsky, 1986, p. 7–8). This distinction has proven very useful to discriminate internal properties of grammars and linguistic features, dependent on external sociolinguistic considerations (Sobin, 1997; Lightfoot, 1999; Lasnik and Sobin, 2000; Madariaga, 2009; etc.).

Among formalists, there is common agreement in that linguistic change is contingent, in the sense that the initial trigger of a shift in grammar is primarily originated in language performance/E-language, which is partly determined by extra-linguistic factors and can change in unpredictable ways (Faarlund, 1990; Lightfoot, 1991, 1999, 2006; Roberts, 2007). The conditions of language transmission can be altered by modifications of the PLD, triggered by external random sociolinguistic factors, phonological erosion, previous unrelated morphosyntactic changes, drops in frequency of the relevant input, etc.

Some authors, however, refine this idea by proposing certain regularities imposed by our Language Acquisition Device (LAD), which can lead learners to acquire a structure in a new way with respect to the previous generation of speakers, thus giving rise to diachronic change. This is depicted by some authors in the form of hierarchies arranging the parametric choices available in acquisition according to more or less marked options. These options determine the probability of a parameter to be set in one way or another and, therefore, the possible ways in which change will most likely take place (Roberts, 2007, p. 267ff, 2012; Biberauer et al., 2010).

Other biases determining, at least partially, language change are considerations of optimality, economy, and a tendency of grammars to become simpler (van Gelderen, 2004). This is in line with Chomsky's (2005) "third factor," which can be defined as those language independent principles of structural architecture, efficiency, and processing that render language as an optimal solution to the interface (phonological and semantic) conditions.

According to these previous notions, the fundamental role of diachronic change as a "language shaper" is then two-fold, as it can affect internal grammars (I-language) or just remain at a "surface" level, modifying the speakers' external productions in E-languages. Here are the views at this respect:

- (i) Diachronic changes affecting I-languages are in the first place related to language acquisition, Chomsky's (2005) "second factor." As we said before, formal approaches to diachrony assume that change takes place between two different generations of speakers during the language acquisition process (cf. an illustrative case study in Duguine and Irurtzun, 2014). With the advent of the minimalist program,

third factor effects are also acknowledged to be implied in diachronic change (Biberauer and Roberts, 2016). Some examples are the Minimax thesis (Chomsky, 2009; Fodor, 2009), according to which parameters must be understood as an optimal solution to the conflict between UG and learnability (“minimize genetic information and optimize the amount of learning”), the role of Feature Economy in grammaticalization processes (van Gelderen, 2004), and the spread of a specific change through different structures or lexical items by Input Generalization (Roberts, 2007; “generalization of the input”).

- (ii) Diachronic changes affecting E-languages, i.e., understood as innovations at an adult age (cf. the concept of “emergent grammars,” as in Hopper, 1987) are mostly disregarded in formal accounts. However, as said before, in the “contingent” type of diachronic generative approaches (Lightfoot, 1991, ff), this kind of innovative or surface modifications of the PLD are acknowledged as potential initial triggers for further changes in I-grammars.

In this respect, a third area must be considered, namely, Externalization processes (Chomsky, 2010; Sigurðsson, 2011), as we usually call the ways of mapping I-features into more external components, e.g., the morphological realization of abstract syntactic features, which will be the central topic in this paper.

All these considerations lead us to ask ourselves about the locus of variation in minimalism. Here we also face different options, which do not necessarily exclude each other: (i) an older idea is the so-called Borer (1984)-Chomsky (2001) Conjecture, that all variation is contained in the Lexicon; (ii) a later refinement of this idea is to admit that the interfaces themselves, in addition to the Lexicon, can also answer for linguistic variation (e.g., Biberauer, 2008, p. 32); and (iii) finally, we observe additional third-factor clustering effects across languages, probably related to the specific ways of mapping syntactic structures into the interfaces (Biberauer et al., 2010; Boeckx, 2011; Roberts, 2012).

In what follows, I will focus on the main goal of this paper, which is to vindicate the role of historical change in formal accounts. This idea does not imply that change has a direct effect on synchronic stages of a language, because we know that speakers do not have access to the I-grammars of previous generations (as represented in Andersen’s, 1973 Abduction principle). But diachronic change definitely can shed light on the ways variation has to be understood, and even on the paths that I-languages follow in order to be configured.

Diachrony interacts with synchronic accounts in different ways, for example, a fundamental reason that led some scholars to revisit cartographic and lexicalist approaches to the synchrony of languages was the need to explain acquisition and change through it (Roberts, 2012). But the study of historical change also helps us understand synchronic language-specific properties and concrete instances of variation (cf. the examples in Lightfoot (in preparation), or even at a methodological level, it can help us decide between two alternative explanations of a synchronic phenomenon (see e.g., Ormazabal and Romero, in preparation).

Following these lines, the case study presented in the following sections constitutes an illustrative example of how diachronic data can clarify the puzzle posited by an instance of variation in a synchronic stage of a language.

## A SYNCHRONIC VARIATION PHENOMENON: CASE ALTERNATION IN RUSSIAN *-sja* VERB OBJECTS

In this section, I provide a synchronic description of our case study. I will focus on the phenomenon of case alternation between genitive and accusative case marking on the object of some medial verbs in Russian, which are virtually all the *-sja* verbs expressing fear or avoidance, as well as the verbs *slušaťsja* “to obey” and *doždat’sja* “to expect.” These verbs display genitive object case marking in standard varieties (1) and accusative case marking in colloquial varieties (2); cf. Peškovskij (2001 [1938], p. 278) and Kryš’ko (1997).

- (1) a. On boitsja ženy. (Standard variant)  
he.NOM fears wife.GEN  
“He is afraid of his wife.”  
b. Devočka vseгда slušaetsja materi.  
girl.NOM always obeys mother.GEN  
“The girl always obeys her mother.”  
c. Inspektor doždalsja kollegi.  
inspector waited colleague.GEN  
“The inspector waited for his colleague.”
- (2) a. On boitsja ženu. (Colloquial variant)  
he.NOM fears wife.ACC  
“He is afraid of his wife.”  
b. Devočka vseгда slušaetsja mamu.  
girl.NOM always obeys mum.ACC  
“The girl always obeys her mom.”  
c. Paren’ doždalsja devušku iz armii.  
young man waited girl.ACC from army  
“The young man waited for his girlfriend from her military service.”

The verbs implied in this alternating pattern are the following (ap. Peškovskij, 2001 [1938], p. 278): (i) all the *-sja* verbs of fear, avoidance, separation: *bojaťsja* “to be afraid,” *stonorit’sja* “to avoid,” *pugaťsja* “to be frightened,” *stydit’sja* “to be ashamed,” *osteregat’sja* “to beware of,” *opasaťsja* “to be afraid, to mistrust,” *strašit’sja* “to dread,” *čuždat’sja* “to keep oneself aloof,” *lišat’sja* “to be deprived,” *stydit’sja* “to be ashamed,” *konfuzit’sja* “to feel ill at ease,” *stesnjaťsja* “to be timid,” etc.; (ii) the weak intensional *-sja* verbs *slušaťsja* “to obey,” and *doždat’sja* “wait, expect” (the only representatives of this kind nowadays).

Timberlake (2004, p. 317) offers a semantic classification of the Russian verbs taking lexical genitive case nowadays (cf. also (Kagan, 2013), for a more detailed semantic account). Those are verbs including the following semantic components<sup>1</sup>:

<sup>1</sup>In the last two groups (types ii and iii), there are two active verbs (without the prefix *-sja*) that display in Modern Russian the alternating genitive vs. accusative

- (i) “Potential,” i.e., “contact is potential but unrealized,” the so-called weak intensional verbs (*iskat* “search for,” *ždat* “await,” *trebovat* “demand,” *želat* “desire,” *xotet* “want”). Almost all of them are active (without the suffix *-sja*), and their alternation pattern is semantically determined, unlike the alternation discussed in this paper. Only the *-sja* forms in this group, *slušaťsja* “obey” and *dožidat’sja* “wait,” follow the pattern (1–2) addressed here, and do not share the semantically determined alternation of their active counterparts, as we will see in the last two sections of this paper.
- (ii) “Avoidance,” whose semantics is described as “possible contact is avoided,” a notion conveyed by the verbs of fear and separation analyzed here.
- (iii) “Tenuous,” defined by Timberlake as “actual contact in the face of possible non-contact” (*dobivat’sja* “achieve, acquire,” *kasat’sja* “touch on”). For the purpose of this paper, I will classify these verbs together with the weak intensional potential verbs, because they share the semantic feature of “potentiality,” and behave in the same way historically.

In this paper, I will leave aside the active weak intensional verbs of the type *iskat* “search,” *ždat* “await,” *trebovat* “demand” (type i) because the distribution of the variants in them is radically different to the one discussed here. Weak intensional active verbs, unlike the verbs under study in this paper, display a clear cut semantic distribution of case marking: roughly, genitive case is used with potential but unreal/“unbounded” objects (usually abstract nouns, but also some concrete but indefinite objects), as in (3a); and accusative case for concrete or real/“bounded” objects, definite or not, as illustrated in (3b) (see Timberlake, 2004; Kagan, 2013).

- (3) a. *My ždēm otveta* / *My trebujem vašego vnimanija.*  
we wait answer.GEN / we demand [your attention].GEN  
“We are waiting for an/the answer/We demand your attention”
- b. *My ždēm žurnal* / *My trebujem naše bljudo.*  
we wait journal.ACC / we demand [our dish].ACC  
“We are waiting for a/ the journal/ We demand our dish.”

The alternations in (3) imply a semantic contrast between genitive and accusative case marking, which is totally absent in the case of the verbs of fear/avoidance, or *slušaťsja* “obey” and *dožidat’sja* “expect,” illustrated in (1) vs. (2). In the latter, much fuzzier factors dealing with declension class, language level, and the speaker’s age are involved, as we will see later on in this paper<sup>2</sup>. The nature of this alternation strongly suggests that we are dealing with a change in progress:

First, there is an undoubtedly high degree not only of interspeaker variation, but also of intraspeaker variation, which

pattern addressed in this paper, as in examples (1–2): *izbegat* “avoid” (type ii), and *dostigat* “reach” (type iii). We will come back to them later on.

<sup>2</sup>Even if some Russian speakers seem to display a semantically-oriented pattern in the distribution of case marking in the verbs of avoidance (at least, with the verb *bojat’sja*), this distribution corresponds to a different semantic feature, namely, the (in)animacy of the NP object (Ora Matushansky p.c.), and not an unreal vs. real feature, as in active weak intensional verbs, as we will see later on.

points to a situation of double coding or, at least, of competing grammars (Kroch, 1989; Yang, 2002), introduced in the previous section.

Second, some authors have observed an increase of the accusative pattern in recent decades, together with a higher frequency of use of the accusative pattern among younger speakers and colloquial registers (Kryśko, 1997; Nessel and Kuznetsova, 2015a).

Finally, this alternation displays the typical “peripheral” properties of certain linguistic phenomena (as described in Baker, 1991; Sobin, 1997; Lasnik and Sobin, 2000; Madariaga, 2009; cf. the distinction between I-level vs. E-level phenomena in the previous section):

- (i) Inconsistent or contradictory productions and intuitions about the variants;
  - (ii) In many cases, free optionality of choice between two variants;
  - (iii) Non-natural late acquisition of one of the variants (learning based on repetition, rules taught at school), or frequent exposition/priming<sup>3</sup>;
  - (iv) The more frequent or colloquial the lexical item at issue, the more often accusative case is used; for example, the occurrences of the accusative variant outnumber those in the genitive case if we perform a simple Google search for the combination *bojat’sja* “be afraid” and *mama* “mom” (colloquial variant), while the percentages differ when we search for the same verb + *mat* “mother,” a less colloquial lexical item, as shown in (4) vs. (5)<sup>4</sup>:
- |        |                             |                      |
|--------|-----------------------------|----------------------|
| (4) a. | <i>bojat’sja mamu.ACC</i>   | 58.8% of occurrences |
| b.     | <i>bojat’sja mamy.GEN</i>   | 41.2% of occurrences |
| (5) a. | <i>bojat’sja mat’.ACC</i>   | 25% of occurrences   |
| b.     | <i>bojat’sja materi.GEN</i> | 75% of occurrences   |
- “Be afraid of one’s mother”

In the following section, I will apply different approaches and hypotheses (non-formal and formal) to this phenomenon

<sup>3</sup>On the one hand, Russian children are taught at school that the “correct” case form for these objects is genitive (at least from Peškovskij, 1918, p. 78). On the other, the role of linguistic priming becomes evident when confronting speakers with a specific language chunk to which they have been frequently exposed. In this case, they will choose the case marking variant corresponding to the “familiar” chunk, even if the alternative form is also in principle available. For example, when speakers are asked to complete the sentence *My boimsja volka i \_\_\_* “we are afraid of the wolf and \_\_\_”, by choosing between the form *sovy.GEN* and *sovu.ACC* “owl,” they unanimously choose the second variant. This is because the accusative variant is precisely the one heard in a very similar sentence of a famous song in the film *Brilliantovaja ruka* (“Diamond arm”).

<sup>4</sup>Accusative object case (as opposed to genitive case) is widely acknowledged to be the colloquial variant in the relevant case alternation, and this Google search is provided just as an illustrative example of this fact. After searching for the relevant combinations with the word order “V + complement” (restricting the search to Russia), cleaning up the hits obtained after each search, and discarding repeated and irrelevant results, we obtain the following figures: the “expected” combination of a colloquial lexical item and accusative case marking (*mamu*) renders 123 occurrences, while *bojat’sja mamy* (colloquial lexical item, genitive case marking) renders 86 occurrences (58.8 and 41.2%, respectively). With the more formal lexical item *mat*, both case marking options diverge much more: *bojat’sja mat*—41 occurrences vs. *bojat’sja materi*—124 occurrences (25 and 75%, respectively).

of variation, and review their advantages and shortcomings. Afterward, I will offer my own proposal, which introduces diachronic data, and show in which way it is more explanatory than the purely synchronic accounts proposed so far.

## SYNCHRONIC APPROACHES ACCORDING TO DIFFERENT LINGUISTIC ORIENTATIONS

### Non-formal Approaches

In this section, I will review three previous studies on this specific topic or in more general, but directly related, phenomena of the Russian language. All these studies have been performed from the perspective of structuralist or functionalist approaches. Albeit there are noticeable differences between them, these approaches appear as just descriptive and, in some cases, also incomplete.

### Decomposition of Grammatical Case

Scholars of structural linguistic orientation explored the possibility of decomposing grammatical case into smaller semantic features. Each grammatical case would be in this way characterized by a group of features that enter into syntagmatic and paradigmatic relations. The presence of common features among different cases would allow for replacing one case with another when they share most features, or extending the uses of a case by addition or loss of the relevant features.

One of the most renowned examples of this system is precisely decomposition of Russian case, proposed by Jakobson (1984 [1936]), and revised later by Franks (1995). These authors do not specifically address the alternation in case marking under study in this paper, but they do examine a related morphological syncretism, namely, the conflation of genitive and accusative morphological cases on animate objects (masculine singular and all plural). This conflation is illustrated in (6b):

- (6) a. Vasilij Ivanovič xorošo znaet *moj gorod*.  
 Vasili Ivanovich well knows [my town].NOM/ACC  
 “Vasili Ivanovich knows my town very well.”  
 (cf. *Ėto moj gorod*.NOM “This is my town”)
- b. Nikolaj Borisovič xorošo znaet *moego zjatja*.  
 Nikolai Borisovich well knows [my son-in-law].  
 GEN/ACC  
 “Nikolai Borisovich knows my son-in-law very well.”  
 (cf. *Ėto rabota moego zjatja*.GEN “This is my son-in-law’s work.”)

According to these authors’ system of case decomposition, genitive case would consist of the features [+oblique, –marginal, –non-ascriptive], while accusative would be defined as [–oblique, –marginal, –non-ascriptive]. In order to obtain (6b), they just erase the distinction between the two forms by the feature [oblique] in an operation that equals both forms with the characterization [–non-ascriptive, –marginal], and renders the morphological syncretism we observe in the language at the synchronic level.

Hypothetically applying the same system at the diachronic level, we could also claim that accusative and genitive morphological cases can alternate by erasing their [oblique]

feature in the relevant context, leaving both forms with equal features ([–non-ascriptive, –marginal]). This operation would render the alternation between genitive and accusative in the complements of the verbs discussed in this paper.

This proposal is very appealing, at least, if we settle for a basic morphological description. Nonetheless, the mechanism of decomposition of grammatical cases—and related morphological operations—is just descriptive, and maybe not so persuasive on the basis of independent evidence. At times, the correspondence of a case to the alleged underlying semantic features is not very informative; for instance, in the specific alternation under study here, the characterization of the genitive and accusative cases does not capture the semantic values of avoidance and potentiality, which clearly differ from the usual values of these two cases in other parts of the grammar.

### Maps of Semantic Notions

Another system inspired by the theory of case decomposition is the more recent idea of representing the semantic values underlying grammatical cases with the help of the so-called “maps of semantic notions.” These maps include the various semantic interpretations of the cases existing in a language or a group(s) of languages, and depict the higher or lower plausibility of syncretism or transfer between cases through the representation of the “geographic” distance between the different values.

The closest study to the phenomenon discussed in this paper is to be found in Clancy (2006), based on Haspelmath (1997). He offers a topology of Slavic case with multidimensional scaling, in which the distance between different functions or semantic values intends to capture frequencies of use, markedness of the variants, and possible changes and syncretisms. Thus, broadening or restricting the “meanings” or semantic values of a specific case should correspond to contiguous or related areas on the semantic map.

Clancy (2006, p. 24) captures the relationships between the semantic values of Slavic case in such a map of semantic notions, depicting the “distance” between those values. Such an approach is interesting from the point of view of case morphology, and in this specific study, it is very detailed. However, as it stands, it is of no use for our analysis, as the semantic notions on Clancy’s map associated with the alternating cases addressed here (“dist from/afraid of,” which stands for the ablative value of the genitive case, including verbs of fear, and “understand,” which stands for regular direct objects) are too far away from each other to accommodate our variants and hypothesize a possible transfer between them. This can be due to the fact that Clancy (2006, p. 25) himself acknowledges that the dataset used for this specific map was a pilot one and thus incomplete, but in any case, such a representation of the semantic values of case is purely descriptive and does not explain the reason for the variation phenomenon under study.

### Paradigms or Construction Networks

Construction networks can be defined as a descriptive tool developed within the Construction Grammar, a functionalist approach to languages. Within this approach, Nessel and Kuznetsova (2015a,b) have addressed the specific phenomenon

discussed in this paper. More specifically, the authors count the occurrences of the accusative and genitive variants associated with several *-sja* verbs in the Russian National Corpus and aim to account for the asymmetries in the use of these alternating variants.

As introduced in the previous section, they find differences depending on the register (whether the utterance is part of a corpus or a spontaneous production), the age of the speaker (accusative more frequent in younger speakers), and the specific lexical item. Interestingly, the verbs more often found in combination with an accusative complement are, according to their search in the National Corpus, the verbs *slušaťsja* “to obey,” *bojaťsja* “to be afraid,” *dožidaťsja* “to wait,” *dostigať* “to reach,” and *izbegať* “to avoid;” there are also differences among them, namely, the verb *slušaťsja* and after it, *dožidaťsja*, are much more frequently combined with accusative than the rest (Nesset and Kuznetsova, 2015a, p. 371–373). We will come back to this fact in the following section.

The main hypothesis in their papers is that a high level of individuation or animacy favors accusative case marking, as happens in other cases of alternation between genitive and accusative case (namely, the syncretism of animate NPs in object position, to which we will return later on). All other factors (declension class of the noun, the intensional or directional semantics of the verb, the “opacity” of the *-sja* suffix) are relegated by them as epiphenomenal.

After describing the conditions for case alternation, Nesset and Kuznetsova accommodate all the relevant variables in a paradigm or construction network, which can be defined as the representation of a specific construction and of some of its subtypes, together with the relevant features, such as markedness of choices, statistical significance, possible diachronic changes, variation, etc. Nesset and Kuznetsova (2015a, p. 388) provide such a construction network for three of the verbs involved in the alternation at issue. But, again, the statistics offered by these authors, as well as the intervening factors, are very illustrative of what is happening in the language, but their network is just descriptive. Another shortcoming of the account is that it overlooks the potential syntactic motivations behind the variants, which will play a fundamental role in the alternation, as we will see in the two final sections.

A final observation regards the imprecise semantic and syntactic characterization of the verbs these authors analyze. If we follow Timberlake’s (2004, p. 317) classification of the Russian verbs taking lexical genitive nowadays (cf. previous section), *slušaťsja*, and *dožidaťsja* are weak intensional verbs (the only “potential” *-sja* verbs), *bojaťsja* and *izbegať* are verbs of avoidance (one medial, the other active), and *dostigať* is a potential verb with active form. The indistinct treatment of all these forms leads the authors to lump together verbs of different syntactic and diachronic behavior.

## Formal Synchronic Approaches

To the best of my knowledge, there are no formal studies on this specific alternation phenomenon of the Russian language, so in

this section I will try to apply more general purely synchronic formal accounts to it.

As a first step, we could just think that the alternation discussed here is not relevant for syntax, i.e., that it arose due to a spontaneous change in the relevant morphological rule instruction; the rule formerly realizing genitive case on the object of these verbs would have just been modified by a new rule specifying that these objects must be accusative.

Of course, this can be true from a strict synchronic point of view, but still some questions remain unanswered: (i) Why does this alternation exist? (ii) Why does it match the distinguishing features of a change in progress? (iii) Why is this alternation not uniform (depending on animacy, declension classes, the presence of *-sja*, etc.). We will answer these questions in the following section.

In a more refined way, we could try to apply to this alternation Sigurðsson (2012) system of regular vs. quirky morphological cases in Icelandic. According to Sigurðsson (2012), the expression of m(orphological)-case corresponds to the Externalization component, i.e., to the different ways of assignment or realization of PF-exponents with respect to underlying syntactic features. Crucially, his system acknowledges the presence of “third factor” properties, namely, markedness of the PF-exponents. In other words, some morphological markers are more or less eligible to encode what is located in the corresponding syntactic heads.

Applying these insights to our alternation in (1–2), partially repeated below for convenience, accusative case in (7b) would be an unmarked variant, while genitive case in (7a) would correspond to a marked (quirky) variant in this specific configuration.

- (7) a. On boitsja ženy. (Standard variant)  
 he.NOM fears wife.GEN  
 “He is afraid of his wife.”  
 b. On boitsja ženu. (Colloquial variant)  
 he.NOM fears wife.ACC  
 “He is afraid of his wife.”

Formalizing this observation, we obtain a characterization of the alternating variants and the shift between them: the marked genitive quirky case variant is represented in (8a), while (8b) stands for unmarked accusative case. The change from genitive into accusative in the relevant contexts would be as in (8c), from marked into unmarked:

- (8) a. Genitive direct objects:  $v^{+++}$   
 b. Accusative direct objects:  $v^*$   
 c. Change:  $v^{+++} > v^*$

This is undoubtedly so at a strict observational level but, as in the previous accounts, it is just descriptive. Sigurðsson (2012) himself acknowledges his system as descriptive, because, he says, it is the only thing we can do when dealing with the Externalization component of the language.

In the rest of the paper, however, I will argue that a more precise (though still formal) account is possible if we pay attention to the diachronic dimension of a phenomenon. More specifically, I will show that what seems like a m-case alternation between genitive and accusative case marking hides in fact two



different structures inherited from a quite complex historical shift that took place in the history of the Russian language several centuries ago.

## A FORMAL DIACHRONIC–SYNCHRONIC ANALYSIS

### The Decline of Bare Lexical Genitive Case in Early Russian

In this section, I will argue for the proposal that the alternation between accusative and genitive case marking related to *-sja* verbs did not originate in a spontaneous change in markedness of the m-cases involved. As an alternative, I will propose that it is, in fact, the last step in a long-term change associated with a global reorganization of case marking in Russian.

First, I will show that the alternation in *-sja* verbs, illustrated in (1–2), is not new to the language, but is rather the replication of a prior change from genitive into accusative object case marking, which had taken place in Middle Russian in active verbs. We will see that we are then dealing with a unique change happening at different moments under distinct structural conditions.

In early Indo-European languages, “bare” grammatical cases (i.e., lacking any overt adposition) were often used as lexical case markers of NPs in a variety of syntactic functions and with diverse semantic values. Later on, we observe a tendency to replace bare case endings by adpositional phrases depending on the language or group of languages (Bauer, 1995; Hewson and Bubenik, 2006)<sup>5</sup>.

This was precisely the case of early Slavic and early Russian. Here, bare cases were regularly used in non-structural positions, namely, encoding “oblique” NPs (adjuncts), and also complements of lexical heads. The examples in (9) illustrate different adjuncts marked with bare lexical cases, alternating already in early Slavic with overt prepositions (Borkovskij, 1978, p. 364ff).

- (9) a. *Otstupi voleju Kyeva.* (*1<sup>st</sup> Novgorod Chronicle*, 36)  
left by-will Kiev.GEN  
“He moved away from Kiev by his own free will.”
- b. *Inii mnozi nesoša i Volodimerju a otudu Kyjevu.* (*Laurentian Chronicle*, 69)  
other many carried him Vladimir.DAT and from-there Kiev.DAT  
“Some of them carried him to the town Vladimir, and from there, to Kiev.”
- c. *Izjaslav sēde Kyevě, Svjatoslavъ Černigově.*  
Iziaslav settled Kiev.LOC Sviatoslav Chernigov.LOC  
(*Laurentian Chronicle*, 55)  
“Iziaslav settled down in Kiev, and Sviatoslav in Chernigov.”

Bare cases, including genitive case, could also encode quirky objects of various types in a regular and much broader way than today.

- (10) a. *I vsego na nemo pytati i bezčinija i*  
and all.GEN to him ask and inactivity.GEN and  
*nevěžestva.*  
ignorance.GEN (*House-Orderer*, 36)  
“He must respond for all, for the things that have not been done and those that he does not know.”
- b. *Zaby inočeskogo obeščanija.*  
forgot [of-monk promise].GEN  
(*Life of Dimitri*, 210b, in Borkovskij, 1978: 353)  
“He ignored the monastic vow.”
- c. *I vsjago togo zapasu ključniku*  
and [all this provision].GEN housekeeper  
*vědati.*  
administrate. (*House-Orderer*, 54)  
“And the housekeeper must take care of all these supplies.”

By that time, the old Slavic case system was undergoing a major reorganization. Bare lexical cases were (i) either replaced by overt PPs (adjuncts), as shown in (11) below, (ii) either reinterpreted as non-lexical or structural cases, (iii) or, some times in the case of dative and genitive cases, lost altogether and replaced with accusative case, as we will see soon.

The replacement of bare lexical adjuncts by PPs was completed by late Old Russian—early Middle Russian. The examples in (11) correspond to a later copy of the same texts from which examples in (9) have been extracted<sup>6</sup>. The only difference between them is the addition of overt prepositions in the case of the later copies:

- (11) a. *Otstupi voleju is Kyeva.*  
left by-will from Kiev.GEN  
(*1<sup>st</sup> Novgorod Chronicle—Commission roll*, 112b)  
“He moved away from Kiev of his own free will.”
- b. *Inii otroci nesoša i k Volodimerju a ottudě k Kyjevu.*  
other fellows carried him to Vladimir.DAT and  
from-there to Kiev.DAT  
(*Radziwill Chronicle*, 69)  
“Some comrades carried him to Vladimir, and from there, to Kiev.”
- c. *Izjaslav sēde v Kyevě, Svjatoslavъ v Chernigov.*  
Iziaslav settled in Kiev .LOC Sviatoslavъ in Chernigov.  
(*Radziwill Chronicle*, 55)  
“Iziaslav settled down in Kiev, and Sviatoslav in Chernigov.”

<sup>5</sup>An illustrative example of this development is the parallel loss of the early IE absolute participial constructions in the different IE groups (*ablativus absolutus* in Latin, absolute genitive in Greek, absolute dative in Slavic, etc.), and their later replacement by different circumstantial complements headed by an overt preposition or conjunction (Bauer, 1995).

<sup>6</sup>The *Commission roll* is a fifteenth-century copy of the *1<sup>st</sup> Novgorod Chronicle* (thirteen to fourteenth century) and *Radziwill Chronicle* is a late fifteenth-century copy of the *Laurentian Chronicle* (fourteenth century).

The replacement of bare genitive lexical case with PPs, together with its reanalysis as a non-lexical case, severely restricted the interpretation of the remaining bare genitive NPs as quirks. The presence of such forms created a “disturbing” piece of evidence in the PLD that learners of Russian received, and tended to be progressively driven out from the language.

At this point, we can already realize the deep historical roots of the synchronic alternation addressed in this paper (1–2). In the following pages, I will show that this alternation was a distant product of this initial reorganization of the Old Russian bare case system. As such, it was ultimately tied to the more general typological change that took place in most early Indo-European languages; namely, a shift from Proto-Indo-European OV into VO word order, i.e., from left-branching into right-branching (Lehmann, 1974; Friedrich, 1975; Watkins, 1976; Luraghi, 1990; Bauer, 1995; cf. discussion in Keydana, in preparation, and Pancheva, 2008 for early Slavic). One of the consequences of this recurrent process in Indo-European implied precisely the replacement of bare lexical cases by PPs headed by overt prepositions (Lehmann, 1993; Bauer, 1995; Hewson and Bubenik, 2006). This is precisely the phenomenon observed in early Slavic, too, as illustrated in (9) vs. (11).

In the rest of this section, I will review the changes related to this general diachronic process, which preceded the alternation in case marking in (1–2). As noted before, this process proceeds in two steps of similar characteristics, but distant in time from each other. First, I will focus on the first step, the change from genitive into accusative case marking associated with regular active verbs, then I will account for the second part of this shift, that affected medial (–*sja*) verbs, and explain why it took place much later than the previous one.

## First Step of the Change: Genitive into Accusative Complements of Active Verbs

### The Loss of Bare Genitive Complements of Active Verbs

Bare genitive case associated with some active verbs was lost as soon as in prehistoric Slavic; other active verbs still display genitive object case in early Slavic. This shift affected several classes of active verbs, including the verbs discussed here, i.e., verbs traditionally classified in Indo-European linguistics as verbs of “separation” (avoidance) and verbs of “desire/achievement and perception” (potential).

#### *Verbs of separation (avoidance)*

They can denote physical or psychological avoidance. Savčenko (2003 [1974]) includes in the first group the Indo-European verbs expressing departure, typically associated with an ablative case that, in the languages with ablative-genitive syncretism, is expressed with genitive case (Greek and Balto-Slavic, including Old Russian cf. 12):

- (12) Se azъ otxožju světa sego.  
 this I leave [world this].GEN  
 (Laurentian Chronicle 54b)  
 “Now I am leaving this world.”

Some verbs maintained this pattern as an archaism until the nineteenth century in Russian, as shown in (13). But their complements were in general reinterpreted as adjuncts quite early, by adding an overt P such as *iz, ot, c* “from” (see example 11a above).

- (13) Nadobno každomu bežat' ètogo Peterburga.  
 need each escape [this Petersburg].GEN  
 (Pisemskij, *Tycjača duš*)  
 “Everyone needs to escape from this Petersburg.”  
 (Nowadays: [pp *iz ètogo Peterburga*] “from this Petersburg”)

Psychological avoidance reflects a metaphoric sense of separation, and corresponds to the psych verbs denoting fear (Schmalstieg, 1983; Šaxmatov, 2001 [1941]). Some of them were active and displayed (ablative-)genitive case assignment in certain Indo-European languages, including early Slavic:

- (14) Jego imene trepetaxu vsja strany.  
 his name.GEN feared all countries  
 (Laurentian Chronicle, 97b)  
 “All the peoples feared his name.”

In Middle Russian, the active verbs denoting fear lost genitive complements and replaced them by overt PPs, nowadays *pered* “before” + NP with instrumental case. This is now the pattern of *robel'* “to hang back,” *trusit'* “to fear, to be in a funk,” *trepetat'* “to tremble, to be afraid,” *drožat'* “to tremble,” etc. (15a). Gorbačevič (1971) reports the last literary archaic uses of bare genitive with active verbs in the nineteenth century (15b)<sup>7</sup>.

- (15) a. Oni drožat' *pered* Bogom  
 They tremble before God  
 (Griboedov, in Peškovskij, 2001 [1938], p. 278)  
 b. Odná liš' ja ljubvi do smerti trušu.  
 one only I love.GEN to death fear  
 “I am the only one who dreads love.”

#### *Desire and achievement verbs and verbs of perception (potential)*

These are verbs denoting “to want,” “to search for,” “to wait,” “to achieve,” and “to hear,” “to see,” “to feel,” etc. These active verbs alternated in early Indo-European languages between genitive and accusative case marking, and most of them changed later into an accusative or PP pattern (Savčenko, 2003 [1974]). This is the case of Old Russian, in which the following active verbs of perception are reported to have displayed an alternating pattern (Borkovskij, 1978, p. 346–347): *čitati* “to read,” *smotriti* “to look,” *slyšati* “to hear,” *slušati* “to listen/to obey,” *viděti* “to see,” *očjutiti* “to feel.”

<sup>7</sup>In Old Church Slavonic, bare genitive alternated with a PP *ot* “from” + genitive case (Borkovskij, 1978, p. 353), a pattern occasionally found in Russian as an archaism until the seventeenth century:

- (i) Xudo tomu žit', kto *ot obuxa* drožit.  
 badly for-this live who from axe-butt.GEN shivers  
 (Proverbs of the seventeenth century, #2486, in Borkovskij, 1978, p. 353)  
 “The one, who fears an axe, has a difficult life.”

- (16) a. I *knižnago poučenyja* slušaita.  
and [bookish teaching].GEN listen  
(*Laurentian Chronicle, 151b*)  
“And you both listen to the teaching of the Bible.”
- b. Ašče kto o(t)ca li *m(a)t(e)re* ne poslušajet.  
if who father.GEN or mother.GEN not hears  
(*Laurentian Chronicle, 18b*)  
“If somebody does not obey his father or mother.”

According to Borkovskij (1978, p. 347), such genitive objects stopped being available in Middle Russian (except for stylistically marked archaisms, which survived much longer in the language), and the verbs became regular transitive verbs taking an accusative object. This shift in case marking is illustrated in example (17), extracted from the late fifteenth-century copy of (16b):

- (17) Ašče kto o(t)ca i *m(a)t(e)rʙ*  
if who father.GEN/ACC and mother.ACC  
ne poslušajet. (*Radziwill Chronicle, 18b*)  
not hears  
“If somebody does not obey his father and mother.”

Together with the verbs of perception, Borkovskij and Kuznecov (2004 [1963], p. 428) classify as genitive object verbs also the verbs denoting desire and achievement; almost all of them were active: *dobyvati* “achieve,” *iskati* “seek,” *ždati* “wait,” *prositi* “ask,” *xotěti* “want,” etc.

- (18) a. Uvėdavъ Onanʙja, xotja emu *dobra*.  
having seen Onanya wanted him good.GEN  
(*1<sup>st</sup> Novgorod Chronicle, 134*)  
“He saw Onanya and wanted to help him  
(lit. wanted the good for him).”  
(*Prayer of Danilo Zatochnik, in*  
Borkovskij and Kuznecov, 2004 [1963], p. 428)
- b. Zane muži *zlata* dobudutъ.  
because men gold.GEN achieve  
“Because men will make money.”

All these verbs are classified nowadays as weak intensional verbs and their historical development was different from the verbs of perception. As noted before, nowadays these verbs maintain the genitive vs. accusative alternation in objects but, unlike the alternation addressed in this paper, it is semantically determined (real/bounded vs. unreal/unbounded feature; cf. examples in (3) above).

### Other verbs

Other Indo-European genitive objects of active verbs, which are relevant for Slavic, are reported in Savčenko (2003 [1974]) and Borkovskij (1978) to have been later reinterpreted as adjuncts (with instrumental case or PP), most of them already in prehistoric times. This was the case of the verbs denoting governing (“to govern,” “to rule”), verbs of “held part” (“to grasp,” “to hold by”), as well as speech verbs (“to say,” “to think,” “to remember”). Others changed into regular accusative objects, with the verbs meaning “taking care,” and sorrow (“to regret,” “to feel sorry”).

### The Rise of New Alternations between the Genitive and Accusative Cases in Non-lexical Positions

In parallel to the loss of bare lexical genitive case, we observe in Middle Russian a significant development of the genitive form as a non-lexical case, which became either reinforced in structural positions previously existing in the language, or spread to new syntactic positions.

The structural positions undergoing the genitive/accusative case alternation are the following: (i) regular animate objects, and (ii) NPs governed by some quantificational or negative head. Let us see some examples of them.

(i) The alternation in regular objects arose in Russian with the extension of the genitive case marker to animate regular objects of the masculine singular declension (*o*-stems) and plural declension (all stems). The process of replacement of the old nominative-accusative form by a genitive form in the relevant animate objects started already in Old Church Slavonic (OCS) (19) and was completed in Early Middle Russian (Krys'ko, 1994). Inanimate objects belonging to these stems remained marked with nominative-accusative case (20):

- (19) a. Ce privěšę č(e)l(ově)kŭ němŭ bėsenŭ.  
here carried [person mute possessed].NOM(/ACC)  
(OCS: *Codex Marianus & Zographensis, Mt. 9:32*)  
“They brought him a mute man who was demon-possessed.”
- b. Privedošę emu č(e)l(ově)ka ġluxa.  
carried him person deaf.GEN(/ACC)  
(OCS: *Liber Sabbae, Mt. 9:32*)  
“They brought him a deaf man.”
- (20) Prinesi *darŭ*. (OCS: *Liber Sabbae, Mt. 8:4*)  
carry gift.NOM(/ACC)  
“Offer him the gift.”

In this way, learners started to be confronted with a consistent alternation between genitive and accusative cases in regular object position.

(ii) Other consistent alternations of a similar nature affected partial or partitive objects, quantified expressions, objects of negated verbs (the so-called genitive of negation), weak intensional verbs, cumulative verbs, and other similar prefixed quantificational verbs (with the prefixes *do-*, *za-*, *pri-*, *na-*; see Straková, 1961).

- (21) Iže ne vŭzimetŭ *kr(ŭ)sta* svoego.  
who not takes [cross his].GEN  
(OCS: *Liber Sabbae, Mt. 10:38*)  
“Whoever does not take up his cross...”

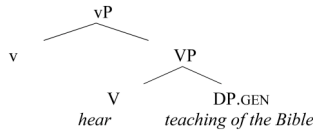
These types of bare genitive case are usually assumed to be licensed by some functional (rather than lexical) head and, therefore, “structurally determined” (Bailyn, 2004; Pereltsvaig, 2006; Kagan, 2013; etc.), thus giving rise to further genitive-accusative alternations in non-lexical positions.

### Formalizing the Change from Genitive into Accusative Objects of Active Verbs

The change experienced by the active verbs reviewed so far can be formalized in the following way: initially, these verbs had the

ability to take genitive objects, as depicted in (22), corresponding to (16a), repeated below:

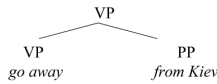
(22) Genitive lexical case pattern (the early Slavic pattern)



(16b) I *knižnago poučēnĭja* slušaita.  
and [bookish teaching].GEN listen  
(*Laurentian Chronicle, 151b*)  
“And you both listen to the teaching of the Bible.”

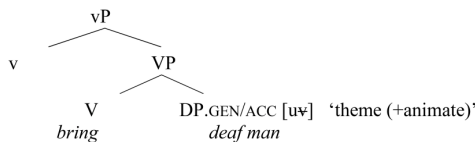
Parallel changes that were taking place in the language at that time (tied to a general typological change in the language) affected mainly two structures: (i) bare genitive adjuncts being replaced by overt PPs (23), corresponding to (11a); and (ii) bare genitive objects alternating with accusative forms in non-lexical (structural) positions (24–25), representing (19b) and (21), respectively.

(23) Replacement of bare genitive adjuncts by PPs



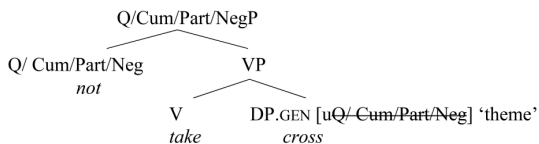
(11a) Otstupi voleju *is* Kyeva.  
left by-will from Kiev.GEN  
(*1<sup>st</sup> Novgorod Chronicle—Commission roll, 112b*)  
“He moved away from Kiev of his own free will.”

(23) Genitive animate NPs in regular object position



(19b) Privedoše emu *č(e)(ovĕ)ka gluxa*.  
carried him [person deaf].GEN(/ACC)  
(OCS: *Liber Sabbae, Mt. 9:32*)  
“They brought him a deaf man.”

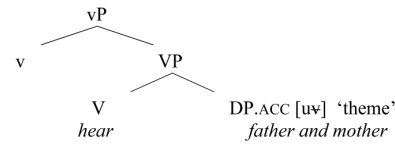
(25) Genitive and accusative NPs alternating in other non-lexical positions



(21) Iže *ne* vŭzimetŭ *kr(ŭ)sta svoego*.  
who not takes [cross his].GEN  
(OCS: *Liber Sabbae, Mt. 10:38*)  
“Whoever does not take up his cross...”

Finally, the result of the change, the regular accusative object case pattern is represented in (26), corresponding to example (17):

(26) Regular accusative object case pattern (the Middle Russian pattern)



(17) Aščē kto o(t)ca i *m(a)t(e)rĕ*  
if who father.GEN/ACC and mother.ACC  
*ne poslušaetĭ.*  
not hears  
“If somebody does not obey his father and mother.”

The change described here can be interpreted according to the concepts outlined in the introductory theoretical section. First, it arises because learners are confronted with innovative pieces of data as part of the Primary Linguistic Data (PLD) they receive, up to a point when their grammar stops converging with the one that generated the relevant input (“discontinuity of transmission between generations”). Language acquisition is therefore a fundamental piece in this process.

Further, as described in the introductory theoretical section, the contingency of grammar change underlies also this specific alternation of the Russian language, as it was determined by the unpredictable alteration of the conditions of the genitive—accusative alternation in other parts of Russian grammar (often related to non-syntactic factors, such as morphological syncretisms).

On the other hand, there seems to exist some bias operating in the previous changes reviewed in this section. There is little doubt that the set of changes associated with a global shift in the basic word order of the language, ultimately responsible for the replacement of bare lexical cases by PPs, is recurrent also in other Indo-European groups of languages (see references above), and seems to respond to some sort of economy or efficiency factor; in this case, to a general tendency to unify the head directionality of the language.

## Second Step of the Change: Genitive into Accusative Complements of *-sja* Verbs

### The Morphosyntactic Development and Formation of *-sja* Verbs

The second phase in the loss of bare genitive case in Russian affected the verbs including a *-sja* suffix, plus the prefixed active verbs *izbegat'* “avoid” and *dostigat'* “reach” to be treated as an exception in the final section. As introduced before, almost all the verbs of fear and avoidance (except *izbegat'* and *dostigat'*) are suffixed *-sja* forms (*bojat'sja* “to be afraid,” *storonit'sja* “to avoid,” *pugat'sja* “to be frightened,” *strašit'sja* “to dread,” *lišat'sja* “to be deprived,” etc.). We include in this section also the weak intensional or “potential” *-sja* verbs *slušaťsja* “to obey” and *dožidat'sja* “to

expect” (the only non-active representatives in their group nowadays).

The history of the *-sja* verbal suffix in Russian is reviewed in detail in Zaliznjak (2008). This suffix has its origin in the clitic *se/sja*, a free morpheme that was in fact the accusative form of the reflexive pronoun (cf. *se* in Romance languages). As such, it could be the object of any active verb regularly taking an accusative complement:

- (27) Na gorě eže sja nune zovet’ Ugor’skoje.  
 on hill which refl.ACC now call.3SG.ACTIVE Ugorskoe  
 (Laurentian Chronicle, 8)  
 “On the hill, which is now called Ugorskoe.”  
 (cf. Spanish = *se llama*/Modern Russian: *nazyvaetsja*.3SG.PASSIVE)

In Old Church Slavonic and Old Russian, unlike in other Indo-European languages (e.g. Ancient Greek), the accusative clitic *se/sja* filled the internal argument position until the sixteenth century (Madariaga, 2010). In Old Russian, these elements could behave as a second-position clitic (28b), or a weak pronoun, usually following the verb (28a), but also following other elements, such as a preposition (Zaliznjak, 2008, p. 36). As we see in example (28b), when the first position in the sentence was occupied by a verb, the clitic could look much like a non-second-position *-sja* to learners (28a), because *sja*, in both patterns, followed the verb:

- (28) a. Knjazja bojsja.  
 prince.GEN fear.refl.  
 (Anthology of 1076, 46)  
 “Be afraid of the prince.”  
 b. Uboiši sja ot lica  
 fear refl. from  
 sill’naago. (Anthology of 1076, 141b)  
 [person strong].GEN  
 “You are afraid of a strong person.”

This kind of input could eventually lead learners to reinterpret the free *sja* as an element associated to a verb. Thus, the free morpheme merged in Old Russian with the verbal form, grammaticalizing later as a verbal suffix (Zaliznjak, 2008). As a final step of this morphological process, the *-sja* suffix underwent phonological reduction into just a palatalized *-s’* in certain environments (in regular conditions, after a final vowel):

- (29) Imeni moego strašilis’.  
 [name my].GEN feared  
 (Tale of Yeruslan Lazarevich, 330, in Borkovskij, 1978, p. 353)

“They were afraid of my name.”

Coming back to the list of the *-sja* verbs alternating between an accusative and genitive pattern in Russian nowadays, we can easily notice that virtually all of them are just the *-sja* “counterpart” of one of the active verbs of avoidance (30a) or potential verbs (30b) reviewed in the previous section. Even nowadays their morphological formation is fully transparent

in most cases: the suffix *-sja/-s’* is just attached to the active form:

- (30) a. Lišat’ “deprive” > lišat’sja “be deprived”  
 Opasat’ “guard” > opasat’sja “mistrust”  
 Pugat’ “frighten” > pugat’sja “fear”  
 Strašit’ “frighten” > strašit’sja “fear”  
 Stranit’ “remove” > storonit’sja “avoid,” etc.  
 b. Slušat’ “listen” > slušat’sja “to obey”  
 Ždat’ “wait” > doždat’sja “expect” (special formation with additional suffix)

Only the verb *bojat’sja* “to be afraid” (Old Church Slavonic *bojati se*) did not correspond to an active verb as such, although prehistoric stages of the language probably displayed an active equivalent as well. Its active counterpart can be traced back to the proto-Slavic form *\*bojati*, not attested as such in historical Slavic, but related to equivalent Sanskrit or Baltic forms. All other verbs of avoidance (30a) display from Middle Russian an active form taking an accusative object, and a *-sja* form taking a genitive object (the one that has recently start to alternate with accusative).

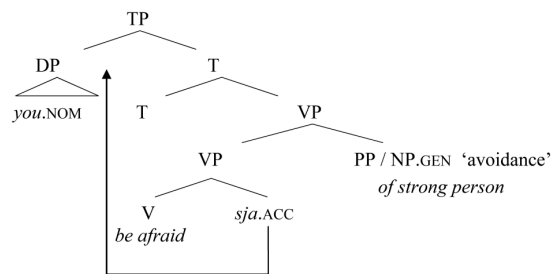
As for the forms in (30b), at the beginning of the twentieth century, by Peškovskij’s (2001 [1938]) time, they had two interesting properties: (i) they were the only members of their lexico-semantic families preserving genitive case (not having changed into an accusative pattern or an alternating pattern determined by the semantics of the object), and (ii) they were the only members of their families with the *-sja* suffix.

Now, after having surveyed all the relevant data, I will propose a formalization of this shift and explain why this change is still taking place nowadays, whereas their active counterparts changed four centuries ago.

### Formalizing the Change from Genitive into Accusative Objects of *-sja* verbs

The original pattern included a free pronominal *sja* element in object position, which could behave as a second-position clitic, as in Old Church Slavonic (cf. example 27). The “avoided” element (i.e., the element causing fear) was associated with the semantics of separation, and marked with genitive (<ablative) case or an overt PP, as usual in most early Indo-European languages. This is illustrated in (31), representing (28b):

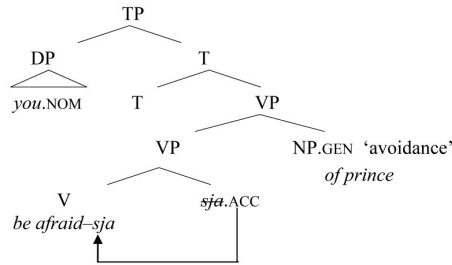
- (31) Old pattern with a free accusative *sja* clitic



- (28b) Uboiši *sja* (ot) *lica sil'naago*.  
 fear refl. (from) [person strong].GEN  
 (Anthology of 1076, 141b)  
 “You are afraid of a strong person.”

In Old Russian, the free morpheme *sja* had the possibility of staying in a lower position and being attached to the verb. This initial short movement prior to reanalysis is represented in (32), corresponding to (28a), in which the pronoun *sja* does not behave as a second-position clitic, but surfaces attached to the verb.

- (32) First morphological incorporation of *-sja* (merge and move)

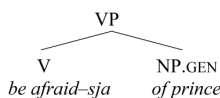


- (28a) Knjazja *boisja*. (Anthology of 1076, 46)  
 prince.GEN fear.refl.  
 “Be afraid of the prince.”

By this time, the complement slot was still occupied by an overt element, which, by virtue of Burzio’s (1986) Generalization, banned accusative assignment to any other possible object. The incorporation of *sja* into the verbal form, which represents the initial step of the change under study here, was completed in Middle Russian (Zaliznjak, 2008, p. 217ff), but it did not automatically convey any further change in the structure at this stage.

Later on, the *sja* element lost its ability to behave as a clitic, and became fully incorporated into the verb in a very common diachronic process classically known as grammaticalization. These kinds of processes have been described in generative accounts as up-the-tree movements, followed by the reanalysis of the element initially moved as base-generated in the landing position (see Roberts and Roussou, 2003). Again, this grammaticalization process was a necessary previous step for later reanalysis, but still did not involve any major change in case assignment. The verbs affected were still acquired as “exceptional” in that their complement was marked with quirky genitive case.

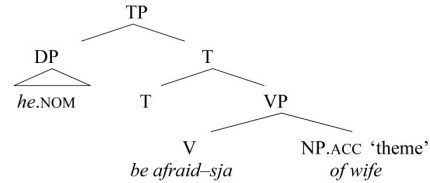
- (33) Reanalysis of *-sja* as directly merged in V



In some speakers, however, at some point in the recent history of Russian, after the morpheme *-sja* started to be base-generated in V, the whole element in this position could start to be perceived as a “deponent” verb. In other words, the complement slot became free, and the historically “disturbing” bare quirky genitive could

finally be reanalyzed as a regular accusative object, merged as a complement of the verb. This is depicted in (34), corresponding to (2a), repeated below:

- (34) Reanalysis of the verbal complement (> shift in case assignment)



- (2a) On *boitsja ženu*.  
 he.NOM fear wife.ACC  
 “He is afraid of his wife.”

The nature of the shift represented in these structures evidences the fact that the ultimate reason for the alternating case patterns with *-sja* verbs, changing recently in Russian, was in fact the reorganization of bare lexical cases four centuries before. In Middle Russian, active verbs taking a genitive complement changed into an accusative pattern (or a semantically determined alternating pattern in the case of weak intensional verbs). But *-sja* verbs behaved in a different way. They preserved a quirky genitive complement longer because at the crucial moment of the reorganization of the bare case system in Russian, they still fell under Burzio’s Generalization; i.e., the complement slot was still filled by the element *sja*, and learners were not able to replace the “disturbing” genitive NP with an accusative NP, as they did in the case of active verbs.

Learners did not have any other option than acquiring the quirky pattern as “exceptional,” as it is still acquired nowadays, i.e., by means of some special morphological rule assigning genitive case to the relevant objects at the Externalization component of the language (Sigurðsson, 2012). This is also in line with the theories about competing grammars coexisting in a single speaker at different linguistic levels, as stated in the introductory theoretical section.

However, after the whole verbal form was reanalyzed as a unique element merged in V, freeing up the complement slot, reanalysis of the “avoided” element as a regular accusative object became available, which is in fact what eventually happened in colloquial language. Again, as explained in the introductory theoretical section, contingent unpredictable conditions determine here the possibility for a syntactic phenomenon to undergo change in a regular way, or the need to “wait” until something else happens in the language, making the conditions for change favorable. The case addressed in this paper is of special interest, because it also confirms the idea that variation can correspond to successive discrete changes spreading further to new syntactic environments.

**On Gradualness and Discreteness of Change**

As said before, the seemingly gradualness of a change in progress can often correspond to a diversity of linguistic environments successively affected by one unique change. Here too, what seems

as a gradual change can be reduced to a series of discrete changes affecting different items or structures at different moments, according to a “third factor” effect, namely, Input Generalization (“maximize available features”); see the introductory theoretical section. If change spreading proceeds in this particular way, we expect the presence of different “splits” between competing variants according to different features, structures, or lexical items. These considerations also apply in our case study.

The major split between the alternating patterns at issue was between active and *-sja* verbs, which have been shown to feature a very clear structural contrast (the availability or not of a free complement slot in the structure).

But other minor splits in this process must also be taken into account, namely, those determined by (in)animacy and declension classes<sup>8</sup>. Some speakers favor accusative case only when the object of the verb *bojat'sja* conveys an animate feature (cf. footnote 2); others reject accusative assignment when the object belongs to III declension class, even if it is animate. These patterns are illustrated in (35):

- (35) Ja bojus' mamu / \*grozu / \*grom  
I fear Mum.ACC.II / storm.ACC.II / thunder.ACC.I  
/\* mat'.  
/mother.ACC.III

Splitting alternating patterns according to some specific semantic feature or morphological class is a recurrent way of pinpointing a change process. In the specific case of Russian, animacy and declension class have played this role before: animacy determine the case patterns for masculine singular I class, and all plural objects (see example 6).

Other speakers, however, have gone further in this process and are able to use accusative case almost regardless of the animacy feature of the object, whether of class I or II (36a-b). Some are tolerant with III declension class objects, too (37a-b)<sup>9</sup>.

- (36) a. U nas inogda daže  
at us sometimes even  
bojatsja bumagu. (Alešin, *Vstreci na grešnoj zemle*)  
fear paper.ACC.II  
“In our country, they are sometimes afraid of some piece of paper.”  
(Google search,  
<https://otvet.mail.ru/question/92115035>)
- b. A elektriki bojatsja grom i  
and electricians fear thunder.ACC.I and  
A elektriki bojatsja grom i  
*molniju?*  
lightning.ACC.II  
“Are electricians afraid of thunder and lightning?”

<sup>8</sup>Class I refers to the 1st declension class (*-o* stems); class II is the second declension class (*-a* stems), and class III stands for the third declension class (*-i* stems).

<sup>9</sup>The verbs most frequently combined with accusative case (*slušaťsja* and *doždat'sja*) accept inanimate and III declension class objects (*slušaťsja mat'/Minfin* “obey mother.ACC.III/Ministry of finances.ACC.I”) more often than *bojat'sja*.

- (39) a. Ja bojus' svoju mat'.  
I fear own mother.ACC.III  
(30 times in a Google search)<sup>10</sup>
- b. Razve možno bojat'sja myš'?  
maybe possible fear mouse.ACC.III  
—udivilsja Birjukov.  
surprised Biriukov  
(Petkevic, *Živye cvety zimoj*)  
“Is it possible to be afraid of a mouse?”  
-asked Biriukov with surprise.”

A final observation on splits in the alternating patterns concerns the spread of the new accusative form according to different lexical items, thus rendering again a succession of discrete changes, as explained in the introductory theoretical section of this paper. As expected, more frequent verbs are more prone to be used with accusative case than others. As noted by Nessel and Kuznetsova (2015a), there are differences between them even in the case of frequently used verbs; namely, *slušaťsja* and *doždat'sja* are changing faster than *bojat'sja*. This correlation is perhaps not random: as shown in (30b) above, both *slušaťsja* and *doždat'sja* have active counterparts, which had changed into an accusative pattern earlier, while *bojat'sja* never had one. This fact suggests that *bojat'sja* was maybe less prone to Input Generalization, and therefore more resistant to change, regardless of its frequent use in the language.

### The Exceptions: The Verbs *dostigat'* “reach” and *izbegat'* “Avoid”

To conclude this section, let us now recall the only active verbs that exceptionally preserved bare genitive objects in the Russian language: *dostigat'* and *izbegat'*. Although they are active, these verbs behave as *-sja* verbs in the sense that they started to change later, and nowadays display, in principle, the alternation addressed in this paper, as shown in (40):

- (40) a. Izbegat' ètoj vetki /% ètu vetku  
avoid [this branch].GEN/ [this branch].ACC  
metro.  
subway  
“Avoid this branch of the subway.”
- b. Dostigat' svoej celi /% svoju cel'.  
reach [own objective].GEN/ [own objective].ACC  
“Reach one's own objective.”

These two verbs are prefixed forms in their basic form, unlike the other active verbs that changed early in the language from genitive into accusative case. On the other hand, they are the only ones lacking a *-sja* counterpart; in other words, like *bojat'sja* (this one lacking an active counterpart), they did not enter the alternation active–medial illustrated in (30) above.

<sup>10</sup>After a search of the exact phrase in (39a), only in Russian pages, I cleaned up the irrelevant and repeated hits, and obtained 30 different occurrences of it. This shows that the use of accusative case with 3rd declension class is not rare at all in the language.

Their special morphology could probably help them to preserve bare genitive case marking as not so “disturbing” to be acquired by learners. In fact, we have independent evidence that bare genitive case was preserved until late Middle Russian, when it was associated with prefixed verbs (Černyx, 1952). The examples in (41) include a verbal prefix, called preverb in the indoeuropeanist tradition, which is identical to the “missing” preposition, making it easily recoverable<sup>11</sup>.

- (41) a. I tu naěxali nas tri tatariny poganye.  
and here flung us.GEN three Tartars evil  
(*Afanasi Nikitin's Journey*, 19)  
“And then three evil Tartars attacked us.”  
(Later: [ppna nas] “toward us”)
- b. Da sluxъ nasъ totъ došelъ.  
and rumor us.GEN this came  
(*Historical acts* 2, 333, in Černyx, 1952, p. 270)  
“This rumor has come to us.”  
(Later: [pp do nas] “to us”)

Likewise, the prefixes in *dostigat'* “reach” and *izbegat'* “avoid,” as well as the lack of voice alternation, could contribute to the longer preservation of the bare genitive complement of the verbs. Again, this development is in line with the introductory notions about successive realizations of one unique change in different morpho-syntactic conditions (see the introductory theoretical section).

## CONCLUSION

In this paper, I have shown the convenience of introducing diachronic analyses into the study of synchronic syntactic phenomena through the practical example of a case alternation in Modern Russian: accusative objects (colloquial pattern) vs. genitive objects (neutral pattern) of the *-sja* verbs denoting avoidance, and the verbs *sluŝat'sja* “obey” and *doŝidat'sja* “expect.”

First, I have reviewed the virtues and shortcomings of previous non-formal accounts about this phenomenon, as well as the potential application of a formal synchronic account to this phenomenon.

Then, I have shown that a formal account including the diachronic dimension is more explanatory. In what sense? The diachronic analysis allows us to realize that the alternation in case marking associated nowadays to *-sja* verbs does not just correspond to a set of morphological rules, but has an additional underlying syntactic explanation. These verbs are now undergoing the same change from genitive into accusative case marking that their active counterparts underwent several

<sup>11</sup>Preverbs of this kind are found in other early Indo-European languages, most famously in Homeric Greek, where some of these adverbial elements were multifunctional; i.e., the same element could behave as a free adverb, a preposition, a postposition or a preverb correlating with a bare lexical case-marked NP, as in the Russian examples in (41).

centuries ago. This change was ultimately tied to the general typological shift experienced in early Slavic, which led to the reorganization of bare lexical cases, especially bare genitive case. The reason why *-sja* verbs started to change later than active verbs is also syntactic: until the sixteenth century, *sja* was an accusative free morpheme, merged as the complement of V; this prevented the change from genitive into accusative marking, which was taking place in active verbs by that time. Merging *sja* with the verb eliminated the obstacle for accusative marking and opened the possibility for these verbs to change following the same path active verbs had undergone some centuries before.

The rest of features characterizing the distribution of the variants according to animacy, declension classes, and lexical items are also accounted for with the help of the diachronic data: splitting the available variants in successive discrete changes according to semantic features (animacy) or morphological features (declension class) is a recurrent phenomenon of pinpointing diachronic processes. On the other hand, a higher difficulty in applying third-factor strategies (Input Generalization) to certain lexical items suggests that these items will be less prone to change, as happens with less frequently used verbs, and also with the verb *bojat'sja* (compared to *sluŝat'sja* and *doŝidat'sja*), because it lacks an active counterpart taking an accusative object. Likewise, differential morphology was probably the reason for slowing down the expected development of the only active verbs (*dostigat'* and *izbegat'*) that preserved the type of case alternation at issue even nowadays.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## FUNDING

The research for this paper has been made possible by the grants from the Spanish Ministry of the Economy and Competitiveness FFI2014-57260-P and FFI2014-53675-P. Support given by the research group on linguistics (UFI11/14) at the University of the Basque Country (UPV/EHU) and the research group on historical linguistics (IT698-13) funded by the Basque Government is also gratefully acknowledged.

## ACKNOWLEDGMENTS

I want to thank the audience at the conferences Formal Approaches to Morphosyntactic Variation, held in Vitoria-Gasteiz 2015, and Formal Approaches to Russian Linguistics, held in Moscow in 2017, especially, to Theresa Biberauer, Pavel Graschenkov, Olga Kagan, Ekaterina Lyutikova, Ora Matushansky, Sergei Tatevosov, and Jenneke van der Wal, as well as the reviewers and the editors from Frontiers, for their insightful observations, comments, and/or discussion on the topic.



## REFERENCES

- Andersen, H. (1973). Abductive and deductive change. *Language* 49, 765–793. doi: 10.2307/412063
- Bailyn, J. (2004). “The case of Q,” in *Formal Approaches to Slavic Linguistics: The Ottawa Meeting, Vol. 12*, eds O. Arnaudova, W. Brown, and M.-L. Rivero (Ann Arbor, MI: Slavic Publications), 1–36.
- Baker, C. L. (1991). The syntax of English *not*: the limits of core grammar. *Linguist. Inq.* 22, 387–429.
- Bauer, B. (1995). *Archaic Syntax in Indo-European: The Spread of Transitivity in Latin and French*. Oxford: Oxford University Press.
- Biberauer, T. (2008). “Introduction,” in *The Limits of Syntactic Variation*, ed T. Biberauer (Amsterdam: John Benjamins), 7–72.
- Biberauer, T., Holmberg, A., Roberts, I., and Sheehan, M. (eds.). (2010). *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge: Cambridge University Press.
- Biberauer, T., and Roberts, I. (2016). “Parameter setting,” in *The Cambridge Handbook of Historical Syntax*, eds A. Ledgeway and I. Roberts (Cambridge: Cambridge University Press), 113–133.
- Boeckx, C. (2011). “Approaching parameters from below,” in *Biolinguistic Approaches to Language Evolution and Variation*, eds A. M. Di Sciullo and C. Boeckx (Oxford: Oxford University Press), 205–221.
- Borer, H. (1984). *Parametric Syntax: Case Studies in Semitic and Romance Languages*. Dordrecht: Foris.
- Borkovskij, V. I. (1978). *Istoričeskaja Grammatika Russkogo Jazyka: Sintaksis - Prostoe Predloženie*. Moskva: Nauka.
- Borkovskij, V. I., and Kuznecov, P. S. (2004 [1963]). *Istoričeskaja Grammatika Russkogo Jazyka*. Moskva: Akademija Nauk SSSR.
- Burzio, L. (1986). *Italian Syntax: A Government-Binding Approach*. Dordrecht: Reidel.
- Černyx, P. J. A. (1952). *Istoričeskaja Grammatika Russkogo Jazyka*. Moskva: Učpedgiz.
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York, NY: Praeger.
- Chomsky, N. (2001). “Derivation by phase,” in *Ken Hale: A Life in Language*, ed M. Kenstowicz (Cambridge, MA: The MIT Press), 1–52.
- Chomsky, N. (2005). Three factors in language design. *Linguist. Inq.* 36, 1–22. doi: 10.1162/0024389052993655
- Chomsky, N. (2009). “Conclusion,” in *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*, eds M. Piattelli-Palmarini, J. Uriagereka, and P. Salaburu (Oxford: Oxford University Press), 379–409.
- Chomsky, N. (2010). “Some simple evo devo theses: how true might they be for language?” in *The Evolution of Human Language*, eds R. K. Larson, V. Déprez, and H. Yamakido (Cambridge: Cambridge University Press), 45–62.
- Clancy, S. J. (2006). The topology of slavic case: semantic maps and multidimensional scaling. *Glossos* 7, 1–28. Available online at: [http://slaviccenters.duke.edu/uploads/media\\_items/7clancy.original.pdf](http://slaviccenters.duke.edu/uploads/media_items/7clancy.original.pdf)
- Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teacher* 35, 125–129. doi: 10.2307/4444260
- Duguine, M., and Irurtzun, A. (2014). From obligatory Wh-movement to optional Wh *in situ* in Labourdin Basque. *Language* 90, e1–e30. doi: 10.1353/lan.2014.0006
- Faarlund, J. T. (1990). *Syntactic Change: Toward a Theory of Historical Syntax*. Berlin: Mouton de Gruyter.
- Fodor, J. D. (2009). “Syntactic acquisition: an evaluation measure after all?” in *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*, eds M. Piattelli-Palmarini, J. Uriagereka, and P. Salaburu (Oxford: Oxford University Press), 256–277.
- Franks, S. (1995). *Parameters of Slavic Morphosyntax*. Oxford: Oxford University Press.
- Friedrich, P. (1975). *Proto-Indo-European Syntax: The Order of Meaningful Elements*. Butte, MT: Montana College of Mineral Sciences.
- Gorbačevič, K. S. (1971). *Izmenenie Norm Russkogo Literaturnogo Jazyka*. Leningrad: Prosveščenie.
- Hale, M. (1998). Diachronic syntax. *Syntax* 1, 1–18. doi: 10.1111/1467-9612.00001
- Haspelmath, M. (1997). *Indefinite Pronouns*. Oxford: Oxford University Press.
- Hewson, J., and Bubenik, V. (2006). *From Case to Adposition: The Development of Configurational Syntax in IE Languages*. Amsterdam: John Benjamins.
- Hopper, P. (1987). “Emergent grammar,” in *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistic Society*, eds J. Aske, N. Berry, L. Michaelis, and H. Filip (Berkeley, CA: Berkeley Linguistic Society), 139–157.
- Jakobson, R. O. (1984 [1936]). “Contribution to the general theory of case: general meanings of the Russian cases,” in *Roman Jakobson. Russian and Slavic Grammar: Studies 1931-1981*, eds L. R. Waugh and M. Halle (Berlin: Gruyter), 59–103.
- Kagan, O. (2013). *Semantics of Genitive Objects in Russian: A Study of Genitive of Negation and Intensional Genitive Case*. Dordrecht: Springer.
- Klima, E. (1965). *Studies in Diachronic Transformational Syntax*. Dissertation, Harvard University, Cambridge, MA.
- Kroch, A. (1989). Reflexes of grammar in patterns of language change. *J. Lang. Variat. Change* 1, 199–244. doi: 10.1017/S0954394500000168
- Krys'ko, V. B. (1994). *Razvitie Kategorii Oduševlennosti v Istorii Russkogo Jazyka*. Moskva: Lyceum.
- Krys'ko, V. B. (1997). *Russkij Istoričeskij Sintaksis. Ob'ekt i Perexodnost'*. Moscow: Indrik.
- Lasnik, H., and Sobin, N. (2000). The who/whom puzzle: on the preservation of an archaic feature. *Nat. Lang. Linguistic Theory* 18, 343–371. doi: 10.1023/A:1006322600501
- Ledgeway, A., and Roberts, I. (2017). *The Cambridge Handbook of Syntactic Change*. Cambridge: Cambridge University Press.
- Lehmann, W. (1974). *Proto-Indo-European Syntax*. Austin: University of Texas Press.
- Lehmann, W. (1993). *Theoretical bases of Indo-European Linguistics*. London; New York, NY: Routledge.
- Lightfoot, D. (1974). “The diachronic analysis of English modals,” in *Proceedings of the First International Conference on Historical Linguistics*, eds J. Anderson and C. Jones (Amsterdam; Oxford: North Holland Publishing Company), 219–249.
- Lightfoot, D. (1979). *Principles of Diachronic Syntax*. Cambridge: Cambridge University Press.
- Lightfoot, D. (1991). *How to Set Parameters: Argument from Language Change*. Cambridge, MA: MIT Press.
- Lightfoot, D. (1999). *The Development of Language: Acquisition, Change and Evolution*. Oxford: Blackwell.
- Lightfoot, D. (2006). *How New Languages Emerge*. Cambridge: Cambridge University Press.
- Longobardi, G. (2001). Formal syntax, diachronic minimalism, and etymology: the history of French *chez'*. *Linguist. Inq.* 32, 275–302. doi: 10.1162/00243890152001771
- Luraghi, S. (1990). *Old Hittite Sentence Structure*. London: Routledge.
- Madariaga, N. (2009). *Grammar Change and the Development of New Case Relations: The Interaction between Core Syntax and the Linguistic Periphery in Old and Present-Day Russian*. Dissertation, University of the Basque Country, Vitoria-Gasteiz.
- Madariaga, N. (2010). The development of Indo-European middle-passive verbs: a case study in Ancient Greek and Old Church Slavonic. *Indogerman. Forschun.* 115, 149–178. doi: 10.1515/9783110222814.1.149
- Madariaga, N. (2012). Formal mismatches and functional advantage in syntactic change: the case of Old and Middle Russian non-verbal predicates. *Diachronica* 29, 231–257. doi: 10.1075/dia.29.2.05mad
- Neset, T., and Kuznetsova, J. (2015a). Constructions and language change: from genitive to accusative objects in Russian. *Diachronica* 32, 365–396. doi: 10.1075/dia.32.3.03nes
- Neset, T., and Kuznetsova, J. (2015b). In which case are Russians afraid?: Bojat'sja with genitive and accusative objects. *J. Slavic Linguistics* 23, 255–283. doi: 10.1353/jsl.2015.0009
- Pancheva, R. (2008). “Head-Directionality of TP in Old Church Slavonic,” in *Formal Approaches to Slavic Linguistics: The Stony Brook Meeting*, eds A. Antonenko, J. Bailyn, and C. Bethin (Ann Arbor, MI: Michigan Slavic Publications), 313–332.
- Pereltsvaig, A. (2006). Small nominals. *Nat. Lang. Linguistic Theory* 24, 433–500. doi: 10.1007/s11049-005-3820-z
- Peškovskij, A. M. (1918). *Škol'naja i Naučnaja Grammatika*. Moskva: Literaturno-izdatel'skij otel Narodnogo Komissariata po Prosveščeniju.

- Peškovskij, A. M. (2001 [1938]). *Russkij Sintaksis v Naučnom Osveščanii*. Moskva: URSS.
- Roberts, I. (2007). *Diachronic Syntax*. Oxford: Oxford University Press.
- Roberts, I. (2012). "Macroparameters and minimalism: a programme for comparative research," in *Parameter Theory and Linguistic Change*, eds C. Galves, S. Cyrino, R. Lopes, F. Sandalo, and J. Avelar (Oxford: Oxford University Press), 320–335.
- Roberts, I., and Roussou, A. (2003). *Syntactic Change. A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press.
- Savčenko, A. N. (2003 [1974]). *Sravnitel'naja Grammatika Indoevropejskix Jazykov*. Moskva: URSS.
- Šaxmatov, A. A., (2001 [1941]). *Sintaksis Russkogo Jazyka*. Moskva: URSS.
- Schmalstieg, W. R. (1983). An Etymology of the word 'to fear' in Indic, Baltic and Slavic. *Lithuanus* 29, 60–65.
- Sigurðsson, H. Á. (2011). Uniformity and diversity: a minimalist perspective. *Linguistic Variat.* 11, 189–222. doi: 10.1075/lv.11.2.03sig
- Sigurðsson, H. Á. (2012). Case variation: viruses and star wars. *Nordic J. Linguist.* 35, 313–342. doi: 10.1017/S033258651300005X
- Sobin, N. (1997). Agreement, default rules and grammatical viruses. *Linguistic Inquiry* 28, 318–343.
- Straková, V. (1961). *Genitivní Slovesná Reke. Kapitoly ze Srovnávací Mluvnice Ruské a České*. Praha: Československá Akademie Věd.
- Timberlake, A. (2004). *A Reference Grammar of Russian*. Cambridge: Cambridge University Press.
- van Gelderen, E. (2004). *Grammaticalization as Economy*. Amsterdam: John Benjamins.
- van Gelderen, E. (2010). "Features in reanalysis and grammaticalization," in *Gradience, Gradualness and Grammaticalization*, eds E. Traugott and G. Trousdale (Amsterdam: John Benjamins), 129–147.
- Watkins, C. (1976). "Towards PIE syntax. Problems and pseudoproblems," in *Papers from the Parasession on Diachronic Syntax*, eds S. Steever, C. Walker, and S. Mufwene (Chicago: Chicago Linguistic Society), 305–326.
- Yang, C. (2002). "Grammar competition and language change," in *Syntactic Effects on Morphological Change*, ed D. Lightfoot (Oxford: Oxford University Press), 367–380.
- Zaliznjak, A. A. (2008). *Drevnerusskie Ėnklitiki*. Moskva: Jazyki slavjanskoj kul'tury.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer RE and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Madariaga. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Syntactic Priming As a Test of Argument Structure: A Self-paced Reading Experiment

Isabel Oltra-Massuet<sup>1,2,3\*</sup>, Victoria Sharpe<sup>4</sup>, Kyriaki Neophytou<sup>2</sup> and Alec Marantz<sup>2,4</sup>

<sup>1</sup> Department of English and German Studies, Rovira i Virgili University, Tarragona, Spain, <sup>2</sup> Neuroscience of Language Lab, NYU Abu Dhabi Research Institute, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates, <sup>3</sup> Serra Hùnter Programme, Generalitat de Catalunya, Barcelona, Spain, <sup>4</sup> Neuroscience of Language Lab, Departments of Linguistics and Psychology, New York University, New York, NY, United States

## OPEN ACCESS

### Edited by:

Ángel J. Gallego,  
Universitat Autònoma de Barcelona,  
Spain

### Reviewed by:

Martin John Pickering,  
University of Edinburgh,  
United Kingdom  
Naoki Fukui,  
Sophia University, Japan

### \*Correspondence:

Isabel Oltra-Massuet  
isabel.oltra@urv.cat

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 03 April 2017

**Accepted:** 17 July 2017

**Published:** 15 August 2017

### Citation:

Oltra-Massuet I, Sharpe V,  
Neophytou K and Marantz A (2017)  
Syntactic Priming As a Test of  
Argument Structure: A Self-paced  
Reading Experiment.  
Front. Psychol. 8:1311.  
doi: 10.3389/fpsyg.2017.01311

Using data from a behavioral structural priming experiment, we test two competing theoretical approaches to argument structure, which attribute different configurations to (in)transitive structures. These approaches make different claims about the relationship between unergatives and transitive structures selecting either a DP complement or a small clause complement in structurally unambiguous sentences, thus making different predictions about priming relations between them. Using statistical tools that combine a factorial 6 × 6 within subjects ANOVA, a mixed effects ANCOVA and a linear mixed effects regression model, we report syntactic priming effects in comprehension, which suggest a stronger predictive contribution of a model that supports an interpretive semantics view of syntax, whereby syntactic structures do not necessarily reflect argument/event structure in semantically unambiguous configurations. They also contribute novel experimental evidence that correlate representational complexity with language processing in the mind and brain. Our study further upholds the validity of combining quantitative methods and theoretical approaches to linguistics for advancing our knowledge of syntactic phenomena.

**Keywords:** structural priming, comprehension, argument structure, unergativity, transitivity

## INTRODUCTION

Research has extensively shown that exposure to a syntactic structure influences to different degrees the way we subsequently process a similar structure in comprehension and production in what has been called *syntactic priming*, *structural priming*, or *structural persistence* (e.g., Bock, 1986; Bock and Loebell, 1990; Bock et al., 1992, 2007; Branigan et al., 1995, 2000; Pickering and Branigan, 1998, 1999; Hare and Goldberg, 1999; Pickering et al., 2002, 2013; Loebell and Bock, 2003; Ferreira and Bock, 2006; Thothathiri and Snedeker, 2006, 2008a,b; Carminati et al., 2008; Hartsuiker et al., 2008; Pickering and Ferreira, 2008; Tooley et al., 2009; Tooley and Traxler, 2010; Segal et al., 2012, 2013; Tooley and Bock, 2014; Traxler et al., 2014; Wittenberg et al., 2014). The main goal of this paper is to use the process of syntactic priming as a behavioral tool to test two competing theoretical approaches to argument structure, namely (i) Hale and Keyser's (1993; 1998; 2002) approach as recently developed in Mateu (2002), Acedo-Matellán (2010), Mateu and Acedo-Matellán (2012), and Acedo-Matellán and Mateu (2013), what we will refer to as the *generative semantics* approach to argument structure, and (ii) Marantz (2005; 2011; 2013), which we will call *interpretive semantics* approach. These two theoretical models illustrate two different views of the syntax-semantics mapping.

Whereas Acedo-Matellán and Mateu’s model operates with semantically unambiguous structures that directly reflect argument/event structure, Marantz’s approach contends that syntax does not necessarily start the derivation with a configuration that transparently represents argument/event structure. The latter thus corresponds to an *interpretive semantics* view of syntax, whereby semantics interprets syntactic structures that do not themselves determine meaning; there might be further semantic readjustments or repair strategies at the interface, similar to those postsyntactic processes found at the morphophonology interface. The former approach, Acedo-Matellán and Mateu’s, is conceived as a *generative semantics* view of syntax in the sense that syntax generates syntactic structures that determine semantic interpretation in a strict one-to-one meaning structure mapping<sup>1</sup>.

Since these theories attribute different syntactic configurations to transitive structures like (2–6) and make different claims about the relationship between transitive structures and unergatives like (1), they make different predictions about priming relations between these sentence types.

former predicts priming between sets (2–3) and (4–5), which are considered to display distinct underlying structures in the latter account.

In order to test these two hypotheses we ran a self-paced reading language comprehension study with 600 subjects over Mechanical Turk. The large number of subjects allows us to model the reading times at the direct object or first PP (Segment 3) and at the second PP (Segment 4) of the same sentences as a function of the structure of the immediately preceding sentence, testing for structural priming within and across sentence types. We conducted a series of statistical analyses and report here the results of two ANCOVAs (Analysis of Covariance) and a linear mixed effects regression analysis on the reading times at Segment 3.

A major headline that can be derived from this study is that we do see syntactic priming effects at all in the context of a behavioral comprehension study on structural priming that uses unmarked unambiguous structures without lexical repetition, i.e., what has been termed lexical boost or lexical enhancement. In addition, our analysis shows a significant effect of the interaction

	Conditions		Segment 1	Segment 2	Segment 3	Segment 4
			NP	V	NP/(PP)	PP
(1)	<b>C1.</b>	Unergative	The dog	barked	in a quiet park	at night.
(2)	<b>C2.</b>	Cognate	The man	dozed	a restful doze	on the train.
(3)	<b>C3.</b>	Creation	The cook	baked	a carrot cake	with spelt flour.
(4)	<b>C4.</b>	Location/Locatum	The girl	saddled	a wild horse	in the farm.
(5)	<b>C5.</b>	Strong transitives	The athlete	ignored	a slight niggle	in his knee.
(6)	<b>C6.</b>	<i>With</i> -Small clause	The worker	loaded	a rail wagon	with hay.

In the generative theory, unergatives (1) are analyzed as derived from transitive configurations, as is standardly assumed since Hale and Keyser (1993), and pattern with cognate object constructions (2) as well as with verbs of creation (3), thus predicting syntactic priming among these sentence types but not between these sets and the remaining types (4–6). The latter are assumed to select for a small clause type complement structure, and are therefore predicted to prime among them in this model. On the other hand, the interpretive account does not predict structural priming between the unergatives (1) and the surface transitives, (2–5), nor between complex complement constructions (6) and the other surface transitive sentences. In this model, sentence types (2–5) are analyzed as transitive configurations, whereas (6) would pattern with double object constructions, as suggested and analyzed in Bruening (“Depictive Secondary Predicates, Light Verb Give, and Theories of Double Object Constructions,” unpublished manuscript, University of Delaware), and unergatives in (1) are not generated as underlying transitive configurations. This means that the interpretive approach does predict some cases of priming that the generative model does not; specifically, the

between conditions—the different types of structures as grouped by the different theories—and priming in trials preceded by two trials of the same category in the interpretive model but not in the generative model, which suggests a potentially stronger predictive contribution of the former model over the latter model. More generally, our experimental study supports the validity of quantitative approaches that combine psycholinguistic methodology with sound theoretical hypotheses about the representation and processing of syntactic phenomena for the study of I-language (Chomsky, 1986, p. 21ff).

### Structural Priming

The novelty of the self-paced reading syntactic priming effects reported in this study is that we do observe syntactic priming effects at all in a study of structural priming in comprehension with unambiguous active sentences without a lexical boost. Let us first summarize relevant aspects of structural priming as a method to test for syntactic structure to set the context of this study.

Our basic initial observation is that the interpretive and the generative models make different predictions with respect to structural priming, the tendency to more quickly repeat or better process a sentence because of its structural similarity to a previously experienced “prime” sentence. Structural or syntactic priming has been studied across modalities, both in

<sup>1</sup>Our use of the labels *interpretive* and *generative semantics* should be strictly understood in the sense just explained in the main text, and not as the two approaches to semantics of the 70s within the theory of transformational grammar, e.g., Katz (1971).

production and comprehension, in behavioral studies. On the one hand, there is consensus that syntactic priming effects in production occur without lexical boost, so that when there is lexical repetition in production, priming effects are boosted or enhanced, e.g., Pickering and Branigan (1998), Segaert et al. (2012), but this is not required to find priming effects. We note here that Pickering and Branigan (1998), in an experiment on completing sentence fragments, report that there is priming without lexical repetition in production only when the target sentence is primed with 2 sentences (but see Mahowald et al., 2016, for a recent meta-analysis that reviews and assesses the current state of knowledge on syntactic priming in language production). On the other hand, most works on syntactic priming in comprehension from different perspectives agree that this is strongly dependent on lexical repetition, e.g., Pickering and Traxler (2004), Branigan et al. (2005), Melinger and Dobel (2005) Arai et al. (2007), Traxler and Tooley (2007, 2008), Tooley et al. (2009), Segaert et al. (2012) and Segaert et al. (2013). That is, exposure to a syntactically related prime sentence leads to a faster reading of a target sentence only if there is lexical overlap of the main verbal head. However, recent studies on structural priming have challenged this view by reporting syntactic priming that is independent from verb repetition in comprehension, specifically Thothathiri and Snedeker (2008a,b), Traxler (2008b), Pickering et al. (2013), and hence also from processing modality, as in Tooley and Bock (2014). We consider here some of the studies on syntactic priming in comprehension in more detail.

Among those that do not observe structural priming in comprehension, Pickering and Traxler (2004) report that there is no priming without lexical boost in this modality on the basis of a reading task with eye tracking recording with sentences containing a reduced relative (cf. Traxler, 2008a). Hence, despite all having the same structure, the sentence in (7a) would prime only (7c), where the main verb is the same, but not (7b).

- (7) a. The man watched by the woman was tall.  
 b. The child cleaned by the girl was covered in chocolate (TARGET-No lexical boost).  
 c. The mouse watched by the cat was hiding under the table (TARGET-Lexical boost).

Arai et al. (2007) report results from two experiments where they investigated whether there is priming during comprehension in ditransitive sentences. Using a visual-world paradigm, whereby participants anticipation of linguistic information was monitored through eye-movement, they observed a priming effect similar to that in production, but only when the verb was repeated between prime and target; that is, the priming effect is completely lexically dependent according to these authors.

Although Segaert et al. (2013) report no differential effects across modalities in an fMRI neuronal study of active and passive sentence comprehension and production, they also point out that there is no syntactic priming among active sentences in the absence of lexical boost of the main verbal head word, even though there is priming among passive structures. Although this is not a behavioral study, but an event-related fMRI study investigating syntactic priming and lexical boost effects on the neuronal activity in brain regions processing syntactic structures

(left IFG and left MTG), it bears directly on our observation that there are priming effects without lexical boost among basic active sentences, even if only after two previous primes. They measure fMRI adaptation of neural activity to repetition of verb-headed syntactic constructions, and report that “there was fMRI adaptation to syntactic repetition when actives had a repeated verb, but no fMRI adaptation to syntactic repetition when actives had a novel verb.” In the case of passives, “there was fMRI adaptation to syntactic repetition both for passives with a repeated verb and for passives with a novel verb.”

More recently, in an eye tracking identification experiment with children, Thothathiri and Snedeker (2008a) find priming effects without lexical repetition in comprehension. As pointed out in Tooley and Traxler (2010), these effects are found in the context of two primed sentences. However, and perhaps more importantly, these same authors further point out that children’s identification involved acting out target sentences with toys, which could potentially be said to invoke some sort of covert production component, in the sense that acting out might involve mechanisms involved in production.

Traxler (2008b) reports the first evidence of between-sentence structural priming in online sentence comprehension without lexical overlap using eye-tracking, where a sentence like (8a), but not sentence (8b), would prime the target sentence (8c), because they both have the same structure, which is different from sentence (8b).

- (8) a. The chemist poured the fluid in the beaker into the flask earlier (PRIME).  
 b. The chemist poured the fluid into the flask earlier (PRIME).  
 c. The vendor tossed the peanuts in the box into the crowd during the game (TARGET).

However, Traxler himself already points out that given that priming here involves adjunct relations and that previous experiments report the impossibility of structural priming of arguments without lexical boost in comprehension, a difference in syntactic processing of arguments vs. adjuncts may be at stake in this case.

Pickering et al. (2013) observe structural priming in both lexically independent and lexically dependent comprehension in a study based on a sentence-picture matching task with ambiguous PP attachment, which can be either high (modifying the verb) or low (modifying the object), as in (9) below. They show that processing is sensitive to the (lexically specific or lexically independent) frequency of an alternative structural analysis, whether through immediate exposure (immediate priming) or via long-term priming, i.e., after some unrelated intervening sentences (persistence of priming).

- (9) a. The policeman is thumping the soldier with the gun (PRIME-lexically independent).  
 b. The policeman is prodding the doctor with the gun (PRIME-lexically dependent).  
 c. The waitress is prodding the clown with the umbrella (TARGET).

Finally, Tooley and Bock (2014) examine structural priming with and without verb repetition in both reading comprehension

and spoken production, using the same prime presentation procedure, the same syntactic structures (reduced relatives, RR, and main clauses, MC), the same sentences, and the same group of participants. They report abstract structural priming in both modalities without significant comprehension vs. production differences in terms of lexical dependency. The first four sentences are primes, while the last two are targets.

- (10) a. The speaker selected by the group gave a great talk (RR-same-PRIME).  
 c. The speaker picked by the group gave a great talk (RR-diff-PRIME).  
 d. The group selected the speaker who gave a great talk (MC-same-PRIME).  
 e. The group picked the speaker who gave a great talk (MC-diff-PRIME).  
 b. The architect selected by the firm had years of experience (RR-TARGET).  
 f. The firm selected the architect who had years of experience (MC-TARGET).

We note that the kinds of stimuli that have been used in structural priming studies are mostly items that require some process of disambiguation. So, what all works have in common is that they observe—or fail to observe—priming effects following syntactically complex material, what Tooley et al. (2009) call “difficult and ambiguous sentence structure,” sentences that are difficult to process and may need re-parsing because up to a specific point they can receive more than one interpretation. Most research, if not all, on structural priming in sentence comprehension is concerned with how subjects resolve syntactic ambiguities or process complex sentences in incremental sentence processing. These include reduced relatives of the type in (10), which have received the most attention to date in comprehension studies, garden-path sentences, like (11), cases of ambiguous high- or low-PP attachment, as in (9), ambiguous double object vs. dative construction, (12), ambiguous datives vs. locatives, (13), or ambiguous locatives vs. passives, (14).

- (11) **Garden-path sentences** (Branigan et al., 1995).  
 a. While the woman was eating the creamy soup went cold.
- (12) **Double Object vs. Dative constructions** (Thothathiri and Snedeker, 2008a,b).  
 a. Give the bird the dog bone.  
 b. Give the bird house to the sheep.
- (13) **Datives vs. Locatives** (Bock and Loebell, 1990).  
 a. The wealthy widow drove her Mercedes to the church (PRIME).  
 b. A rock climber sold some cocaine to an undercover agent (TARGET).
- (14) **Locatives vs. Passives** (Bock and Loebell, 1990).  
 a. The foreigner was loitering by the broken traffic light (PRIME).

- b. The referee was punched by one of the fans (TARGET).

The case in (15) is different. Segaert et al. (2012, 2013) observe that whereas passive structures prime passives, active primes do not have any effect, which seems to argue for a higher priming power of marked structures like passive over unmarked active sentences.

- (15) **Actives vs. Passives** (Segaert et al., 2012, 2013).

- a. The woman serves the man.  
 b. The man is served by the woman.

Even though ambiguity plays no role in this last case, it confirms, then, that structural priming studies share complexity of processing as a fundamental premise to test their priming hypotheses.

One of the main goals of our experimental study is to show that there is priming in unmarked non-incrementally disambiguating contexts, i.e., in simple active sentences.

## Persistence of Priming

Another important feature worth bearing in mind is the persistence of priming, since the design of our experiment is a cumulative running priming paradigm where each target sentence also serves as a prime sentence for the next target sentence. This raises the question of the effects of short-term priming vs. long-term priming. Syntactic priming that persists across unrelated intervening sentences has generally been observed in production (e.g., Bock and Griffin, 2000). All the work we have found on long-term priming in comprehension seems to involve the repetition of the verbal head. On the one hand, Hartsuiker et al. (2008), using a picture description task, show that an enhanced priming effect due to lexical boost does not persist across any number of intervening structures in production. On the other hand, Carminati et al. (2008), using an eye tracking identification task, report that lexically dependent syntactic priming effects persist across two intervening sentences in comprehension. Also in Pickering et al. (2013), it is shown that priming persists with lexical repetition over intervening material in comprehension. More recently, Tooley et al. (2014) have observed structural persistence between prime and target across unrelated filler sentences in sentence priming both in production and comprehension on the basis of event-related potentials (ERP) and eye tracking measures. In their experiments, they use prime sentences containing a reduced relative clause, i.e., a complex and ambiguous structure. We do not consider persistence of priming across intervening sentences in our study, since it is still to be determined whether there is priming at all in comprehension, and whether this persists across intervening material when priming with unmarked unambiguous structures.

## Two Theories of Argument Structure

As pointed out in Marantz (2005) (see also Poeppel and Embick, 2005), generative grammar can and should serve as a source of theoretical hypotheses about the representation of language in the mind and brain and how this is processed, to be formally assessed through standard experimental methods. In this paper

we take two competing theories of argument structure, (i) Acedo-Matellán (2010); Mateu and Acedo-Matellán (2012), and Acedo-Matellán and Mateu (2013), and (ii) Marantz (2005; 2011; 2013) and test their claims and predictions with respect to the representation and processing of syntactic argument structure. Both theories are framed within Chomsky's Minimalist Program, and they both adopt a neoconstructionist view of syntax, whereby argument structure is not lexically projected<sup>2</sup> but created in the syntax by the computational system, a single generative engine for all structure building where minimal units of syntactico-semantic features are combined through the operation of *merge* to create hierarchical syntactic structures that will then receive a semantic and phonological interpretation. Such a basic assumption makes them especially suited for the application of the standard psycholinguistic methodology that correlates representational complexity with computational complexity in the brain, i.e., the hypothesis that "the longer and more complex the linguistic computations necessary to generate the representation—the longer it should take for a subject to perform any task involving the representation" (Marantz, 2005, p. 439). That means that specific differences such as how to merge a root in syntax, whether as a complement or as an adjunct (Acedo-Matellán, 2014), can be reduced to differences in surface syntactic representations of verbal argument structure in the sentences under study<sup>3</sup>. As pointed out in the literature on structural priming, syntactic priming is sensitive or attributable to surface structure, not to abstract structure (e.g., Bock et al., 1992; Pickering et al., 2002; Pickering and Ferreira, 2008; Wittenberg et al., 2014). In this respect, it is worth emphasizing that in both models, the proposed structures are surface structures<sup>4</sup>.

Such fundamental assumptions and similarities between both theories allow us to make use of structural priming as a tool to test a variety of unergative and transitive configurations by measuring reading times at the point where both theories differ in the representation of those syntactic structures, namely between the verb and the first complement (Segment 3).

Before going into the details of our experimental study, the remainder of this section briefly reviews the main claims about the syntax of transitive and intransitive predicates made in the two theoretical models of argument/event structure under study and their predictions with respect to structural priming.

<sup>2</sup>The possibility of having priming of lexical argument structure of the type proposed in Trueswell and Kim (1998), rather than syntactic priming, is thus excluded in these models.

<sup>3</sup>Other theories of argument structure, such as those within monostratal theories of syntax like Generalized Phrase Structure Grammar or Role and Reference Grammar, do not share the same basic assumptions with respect to syntactic argument structure building, and could not be easily integrated within our experimental study.

<sup>4</sup>In Hale and Keyser (1993, 1998, 2002), syntactic configurations corresponded to pre-syntactic abstract structures, i.e., generated at *l(exical)-syntax*, prior to *s(yntactic)-syntax*. However, they are analyzed as surface structures generated in syntax proper in Acedo-Matellán (2010) and Acedo-Matellán (2010), Acedo-Matellán and Mateu (2013), as explicitly stated in e.g., Acedo-Matellán (2010, p. 52).

### The Generative Approach to Argument Structure: Hale and Keyser (1993, 1998, 2002); Acedo-Matellán (2010); Mateu and Acedo-Matellán (2012); Acedo-Matellán and Mateu (2013)

In this strict configurational model of argument structure, compositional semantics is directly read off the syntactic structure. Leaving aside unaccusative structures, the configurations advanced in Acedo-Matellán and Mateu's work for the sentence types under study are (16–18).

In the case of unergatives in (16), already since Hale and Keyser's work, the root,  $\sqrt{\quad}$ , is generally understood as merged in the complement position of a functional head  $v$ . The phonological material of the root is then incorporated into this null verbal head  $v$ . As pointed out in Acedo-Matellán (2010, pp. 53–54), "the structure of unergative verbs as transitives is forced by the properties of the system: it is not possible for a functional head to project a specifier without projecting any complement, since the first DP/root merged with a functional head must be its complement." This also includes cognate object constructions, which would also have a configuration as in (16b).

- (16) Unergative, cognate object, and transitive verbs of creation and consumption.
- Sue danced.  
[<sub>VP</sub> [<sub>DP</sub> Sue] [<sub>v'</sub>  $v$   $\sqrt{\text{DANCE}}$ ]].
  - Sue did a dance.  
[<sub>VP</sub> [<sub>DP</sub> Sue] [<sub>v'</sub>  $v$  [<sub>DP</sub> a dance]]].

The syntactic structure in (16), [ $v$  + DP/ $\sqrt{\quad}$ ], is thus the configuration attributed to unergatives (C1), cognate object structures (C2) and creation verbs (C3) in this model.

On the other hand, (a)telic transitive events, exemplified in (17–18), are all derived from a small clause predicate configuration—whether simple, with a single PlaceP, or complex, with a Place P c-commanded by a PathP (cf. Jackendoff, 1973; Cinque and Rizzi, 2010). In both cases, there is a Figure that moves with respect to a potential Ground (Talmy, 1975). A single relational functional (prepositional) head  $p$  (Hale and Keyser's central coincidence P), interpreted as a PlaceP, introduces a Figure-Ground configuration that establishes a location or state. If further c-commanded by a second head  $p$  (Hale and Keyser's terminal coincidence P), this is interpreted as a PathP and introduces a transition that encodes the change. As with unergatives, the root is merged in complement position of the lower null functional  $p$  head and the phonological material of the root is then successively merged up to the null verbal head.

- (17) Atelic transitive events.
- Sue pushed the car.  
[<sub>VP</sub> [<sub>DP</sub> Sue] [<sub>v'</sub>  $v$  [<sub>PlaceP</sub> [<sub>DP</sub> the car] [<sub>Place'</sub> Place  $\sqrt{\text{PUSH}}$ ]]]].
  - Sue lengthened the rope (for 5 min).  
[<sub>VP</sub> [<sub>DP</sub> Sue] [<sub>v'</sub>  $v$  (=en) [<sub>PlaceP</sub> [<sub>DP</sub> the rope] [<sub>Place'</sub> Place  $\sqrt{\text{LONG}}$ ]]]].
- (18) Transitive events of change of state or location.
- The strong winds cleared the sky.

- [<sub>VP</sub> [<sub>DP</sub> The strong winds] [<sub>v</sub> v [<sub>PathP</sub> [<sub>DP</sub> the sky] [<sub>Path</sub> Path [<sub>PlaceP</sub> [<sub>DP</sub> the sky] [<sub>Place</sub> Place √CLEAR]]]]].
- b. Sue shelved the books.  
[<sub>VP</sub> [<sub>DP</sub>Sue] [<sub>v</sub> v [<sub>PathP</sub> [<sub>DP</sub> the books] [<sub>Path</sub> Path [<sub>PlaceP</sub> [<sub>DP</sub> the books] [<sub>Place</sub> Place √SHELF]]]]].

Thus, all telic and atelic structures are assigned a syntactic configuration where a null verbal head *v* takes a small clause structure, a *pP*, in complement position, which will be a *PlaceP* for atelic predicates or a *PathP* with telic predicates, i.e., a small clause configuration in both cases. This is the structure attributed to location/locatum predicates (C4), like *They saddled the horse*, and strong transitive predicates (C5), like *He ignored the truth*, despite their surface appearance as simple transitive sentences. *With*-small clauses (C6) would also have this syntactic representation, the difference being that the preposition in this case is phonologically realized, not null, and there is therefore no conflation.

**The Interpretive Approach to Argument Structure: Marantz (2005, 2011, 2013)**

On the basis of empirical evidence based on the syntax and semantics of *re*-affixation, the interpretation of roots in denominal verbs and restrictions on the interpretation of verbal compounds, Marantz (2011) argues that roots cannot merge as complements of a null functional head, as in Acedo-Matellán and Mateu’s structures (16–18), but must merge as event modifiers, i.e., as adjuncts.

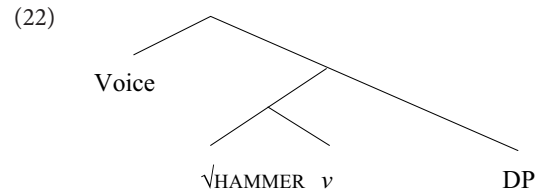
We review here the empirical argument based on *re*-affixation. *Re*- prefixation distinguishes between unergative and transitive structures, as in (19), and between verbs selecting a single direct object and those that take two in a small clause configuration, as in (20). On the one hand, restitutive *re*- is restricted to verbs with an underlying direct object (Horn’s, 1980; generalization); on the other hand, that direct object must be the sole obligatory constituent within the VP (Wechsler’s, 1989 generalization). Hence, the ungrammaticality of (19b) must thus be due to the absence of an underlying object, whereas the grammaticality of (20c) argues against its alleged status as a small clause predicate.

- (19) a. John danced.  
b. \*John re-danced.  
c. John re-danced a dance first performed by his distant ancestors.
- (20) a. John put the display \*(on the table).  
b. \*John re-put the display on the table.  
c. John re-shelved the books.

This means that the root *dance* cannot have been generated in the complement position of the verbal head *v*, because there is no direct object present that *re*- can target in (19a); likewise, *shelve* cannot have a small clause configuration, as proposed in Hale and Keyser and Acedo-Matellán and Mateu, since it does take a direct object that *re*- can target. Marantz concludes that unergatives are plain intransitive predicates, whereas sentence types C2-C5 contain plain transitive predicates, i.e., verbs of creation and incremental themes, unergative verbs

with a cognate object, strong transitives, as well as atelic and telic transitives—which includes location and locatum predicates. The structure is illustrated in (22) for a predicate like *hammer the nail* in (21); the root *hammer* modifies the event introduced by *v* in (22), which selects an internal argument DP, *the nail*.

(21) hammer the nail.



**Predictions of Each Model**

Since these two theoretical approaches to argument structure attribute different configurations to (in)transitive structures, they make different claims about the relationship between them, and therefore make different predictions about priming relations between these sentence types.

In the *generative* model, unergative verbs (C1) share their transitive syntactic configuration with cognate objects (C2) and verbs of creation (C3), whereas location/locatum structures (C4) and strong transitives (C5) pattern with predicates containing a *with*-small clause (C6). In the *interpretive* model, however, the grouping is organized in three different sets, where cognates (C2), creation verbs (C3), location/locatum (C4) and strong transitives (C5) pattern together in a group separate from unergatives (C1) and small clauses (C6). These differences are represented in **Table 1**, where we have identified each sentence type as a priming condition, C1–C6.

Given the 6 sentence types we have singled out and the different structural configurations they are assigned in each theory, we identified the divergent individual priming predictions by sentence type made by each model. These are summarized in **Table 2**. Here we leave aside default identity

**TABLE 1 |** Priming conditions, sentence types and groupings by theory.

VERB TYPE	Generative	Interpretive
C1 UNERGATIVE VERB The dog barked in quiet parks at night.		v
C2 COGNATE OBJECT The man dozed a restful doze on the train.	v + √/DP	
C3 CREATION He baked a delicious cake with spelt flour.		v + √/DP
C4 LOCATION/LOCATUM They saddled a wild horse in the farm.		
C5 STRONG TRANSITIVES He ignored a slight niggle in his knee.	v + SC	
C6 WITH-SMALL CLAUSE They sprayed a cookie sheet with vegetable oil.		v + SC



**TABLE 2 |** Priming relations-predictions of each model by individual sentence types.

	PRIME	TARGET	Generative	Interpretive
C1>C2	UNERGATIVE	COGNATE	✓	✗
C1>C3	UNERGATIVE	CREATION	✓	✗
C2>C1	COGNATE	UNERGATIVE	✓	✗
C2>C4	COGNATE	LOCATION/LOCATUM	✗	✓
C2>C5	COGNATE	STRONG TRANSITIVE	✗	✓
C3>C1	CREATION	UNERGATIVE	✓	✗
C3>C4	CREATION	LOCATION/LOCATUM	✗	✓
C3>C5	CREATION	STRONG TRANSITIVE	✗	✓
C4>C2	LOCATION/LOCATUM	COGNATE	✗	✓
C4>C3	LOCATION/LOCATUM	CREATION	✗	✓
C4>C6	LOCATION/LOCATUM	WITH-SMALL CLAUSE	✓	✗
C5>C2	STRONG TRANSITIVE	COGNATE	✗	✓
C5>C3	STRONG TRANSITIVE	CREATION	✗	✓
C5>C6	STRONG TRANSITIVE	WITH-SMALL CLAUSE	✓	✗
C6>C4	WITH-SMALL CLAUSE	LOCATION/LOCATUM	✓	✗
C6>C5	WITH-SMALL CLAUSE	STRONG TRANSITIVE	✓	✗

**TABLE 3 |** Priming relations-predictions of each model by sentence groupings.

	PRIME/TARGET AMONG THEMSELVES	Generative	Interpretive
C1-C2-C3	UNERGATIVE-COGNATE-CREATION	✓	✗
C4-C5-C6	LOCATION/LOCATUM-STRONG TRANSITIVES-WITH SMALL CLAUSE	✓	✗
C2-C3-C4-C5	COGNATE-CREATION-LOCATION/LOCATUM-STRONG TRANSITIVES	✗	✓

priming for each individual condition, as well as predictions shared by both models, e.g., priming between C2 and C3. Thus, under structural priming conditions, we would mainly expect faster reading times for the first constituent after the main verb—Segment 3—if the sentence involved follows one (or two) sentences of the same structural type. This is the place where the two models structurally differ with respect to the type of complement, a DP or a small clause. That is priming effects would show up as an effect of the primed/unprimed variable of interest—indicated as checks or crosses on **Table 2**—based on each theoretical model.

When considered in terms of the structural groupings and the predictions of each theory with respect to structural priming effects within and across sentence types, the differences between the two theoretical models are summarized in **Table 3**. Thus, the *generative* model predicts priming (i) among unergatives, cognate object constructions and creation verbs and (ii) among location/locatum structures, strong transitives, and structures containing a *with*-small clause. However, the *interpretive* theory predicts priming only (i) among cognate object structures, creation verbs, location/locatum predicates and strong transitives, while (ii) unergatives, and (iii) *with*-small clauses would not show priming effects in prime/target interactions with other sentence types.

In the statistical analyses we discuss in the following sections we analyze the priming relations predicted in **Table 3**, rather than the individual priming relations listed in **Table 2**.

## MATERIALS AND METHODS

### Participants

We distributed our study via Amazon Mechanical Turk to 600 subjects, from which we obtained 460 full datasets<sup>5</sup>. We restricted this to participants from the U.S and those that had a 95% or greater HIT acceptance rate. Data was processed before starting the analysis, and all non-native English speakers were excluded, together with those that spoke more than one language, English, leaving only 390 monolingual native English participants. Within these 390 datasets, only 375 were unique participants; hence, duplicate participants were excluded as well, and only their first set of data was taken. Finally, out of the remaining 375 datasets, 20 were excluded, i.e., about 3%, which correspond to those that had less than 70% overall accuracy on the questions. This resulted in a total number of 355 participants in the included data set, from which 123 male, 166 female, 66 declined to provide demographic information; mean age was 41.38 (*SD* = 12.92).

This study was carried out in accordance with the recommendations of the NYU University’s Institutional Review Board (IRB). All subjects gave written informed consent before beginning the experiment in accordance with the Declaration of Helsinki.

<sup>5</sup>That means that either (i) some subjects completed HITs without doing the experiment, or (ii) some of the datasets did not get saved on the Ibex server.

## Materials

The experimental stimuli consisted of a total of 144 sentences, divided into the 6 different types of structures exemplified in **Table 1** (6 types × 24 sentences = 144).

We have been exhaustive in including as many conditions as structural differences there are between the two models. Thus, sentence types were selected on the basis of the basic syntactic structures attributed to them in the two models under study. Structuring them into 6 types covers all (in)transitive and small clause patterns. For instance, even though creation verbs (C3) and strong transitives (C5) surface as transitives, they have the same structure in the *interpretive* model, but they are attributed different syntactic structures in the *generative* approach, already so since Hale and Keyser’s (1993) seminal work. Therefore, the two models predict different priming effects between these conditions as well as in their interaction with the rest of conditions. To wit, as shown in **Table 2**, whereas creation verbs (C3) and strong transitives (C5) are predicted to prime each other in the *interpretive* model, they are not in the *generative* model. Likewise, although creation verbs (C3) would prime unergatives (C1) in the *generative* framework, strong transitives (5) do not prime them; neither creation verbs (C3) nor strong transitives (C5) would prime unergatives in the *interpretive* theory.

Specific verbs were selected on the basis of the frequency rates of the syntactic patterns they may appear in as reported in the VALEX subcategorization corpus (Korhonen et al., 2006). Specifically, unergative verbs (C1/C2) were chosen on the basis of their low frame frequency with NP complements (frequency lower than 0.15). Creation verbs, Location/Locatum predicates<sup>6</sup>,

We designed a structural priming experiment with six different conditions on sentence structure to run a self-paced reading language comprehension study over Mechanical Turk. Structural priming was tested within and across sentence types using a priming paradigm where each target item also served as a prime sentence for the next target item. In addition, we included an attention task and control condition, which was organized as follows. Every set of 24 sentences had 6 sentences linked to a two-choice comprehension question of the type in (23–24), with a total of 36 questions. These questions served the double function of being an attention task and a control condition to obtain additional reading times from the same prime/target sentences in non-primed contexts.

(23) He dodged a corporate tax in the UK.

(24) Did he evade the tax or pay it?

1. evade it
2. pay it

A complete list of 144 sentences by condition with the corresponding 36 attention tasks linked to 6 individual sentences on each condition is provided as Supplementary Material.

## Procedure–Study Implementation

Sentences of each condition (24 × 6 = 144) were separated into 4 segments, Segment 1–Segment 4: Subject (Segment 1), Verb (Segment 2), First Complement (Segment 3), Second Complement (Segment 4).

	Conditions		Segment 1 NP	Segment 2 V	Segment 3 NP/(PP)	Segment 4 PP
(25)	<b>C1.</b>	Unergative	The dog	barked	in a quiet park	at night.
(26)	<b>C2.</b>	Cognate	The man	dozed	a restful doze	on the train.
(27)	<b>C3.</b>	Creation	The cook	baked	a carrot cake	with spelt flour.
(28)	<b>C4.</b>	Location/Locatum	The girl	saddled	a wild horse	in the farm.
(29)	<b>C5.</b>	Strong transitives	The athlete	ignored	a slight niggles	in his knee.
(30)	<b>C6.</b>	<i>With</i> -Small clause	The worker	loaded	a rail wagon	with hay.

and *With*-Small clause structures were selected from among those with the highest frame frequency rate in the corresponding structure. Strong transitives were chosen on the basis of their high frame frequency with NP complements (frequency higher than 0.83). In addition, combinations of V+N and A+N were checked against the Corpus of Contemporary American English’s (COCA) lexical collocations (Davies, 2008). We also took the definiteness of the NP in Segment 3 into account, as it has been shown that it plays a role in language processing (e.g., Warren and Gibson, 2002). All sentences were further tested against native speaker judgments to confirm naturalness.

We used a running priming paradigm, so that each target sentence served as the prime sentence for the next target item (e.g., Segaert et al., 2012, 2013). Sentences were organized in 3 blocks of trials, with 6 block orderings. Trials were randomized within blocks, so that the conditions followed each other in a random order that was different for each participant. One in every 6 trials was followed by a two-choice comprehension question.

The study was created in Ixet Farm. Participants were shown instructions and they completed a short practice round before the actual experiment started. As a self-paced reading experiment, participants determined the rate at which sentential segments were presented on the monitor by pressing a button, which allowed us to measure reading times at each segment. Each segment was presented sequentially in the center of the screen with 400 ms between each sentence.

<sup>6</sup>Location and Locatum verbs were first selected from among Clark and Clark’s (1979, pp. 769–773) classification. There are 12 sentences with Location verbs and 12 with Locatum verbs.

## Preprocessing and Statistical Model Analyses

Before running the statistical analysis, we calculated the average reading time for each participant, for each segment and then we removed outliers based on this. That is, we excluded trials with reaction times greater than 2 standard deviations from the participant's respective mean. We also excluded the first trial of each block, because it did not fit into either our primed conditions or the unprimed question conditions.

Based on our basic hypothesis that differences in priming effects are expected at the point where both models differ structurally, we decided to focus on the reading times of Segment 3, the first constituent after the verb. Depending on the model, in that position we have a DP complement, a small clause complement or an adverbial. The validity of this hypothesis seemed to be confirmed by the results from a preliminary Analysis of Variance (ANOVA) ( $6 \times 6$  within subjects; Factors: condition + previous\_condition) and visual inspection of the plots, as these differences between sentence types seemed most pronounced in this segment. The controlled analyses that follow all have Segment 3 reading time as the outcome/response variable.

The main analysis of the data was conducted through two different forms of Analyses of Covariance (ANCOVA), for single priming trials and for double priming trials, respectively, and a linear mixed effects regression model, to tailor the two different theories. The use of statistical control allowed us to measure different variables in addition to the independent variables of interest and to control for unexplained variation. For instance, the ANCOVA allows us to have factors as predictors rather than just continuous variables as in a linear regression model.

In both the ANCOVA and the Linear mixed effects regression analysis, the null hypothesis is that all coefficients equal 0. That means that none of the independent variables have any relationship with or effect on the dependent variable, i.e., on the reading time of Segment 3. The alternative hypotheses are that at least one independent variable is predictive of the dependent variable; thus, at least one coefficient does not equal 0.

The statistical analysis was performed using the R Core Team (2015) software program with packages lme4 and lmerTest.

### Single Priming Analysis: ANCOVA 1.0

Following standard procedures, in order to control or minimize the effects of extraneous sources of variance, we included the nuisance variables listed in (31) as covariates. Note that trial order was included as it may account for some of the variance in reading time, e.g., participants may get faster as they proceed or they may change their strategy later in the experiment. Random intercepts by subject and by item were also included in the models.

- (31) a. trial order  
 b. verb frequency  
 c. reading times (RT) of previous segment  
 d. RT of same segment in previous trial

We coded two variables,  $V_I$ - $V_G$ , on the basis of how each theory, interpretive and generative, groups the various conditions, (32).

- (32) a.  $V_I$ -Sentence Types: Unergatives (C1), DP/Root (C2-C5), Small Clause (C6).  
 b.  $V_G$ -Sentence Types: DP/Root (C1-C3), Small Clause (C4-C6).

That means that we took the 6 initial conditions, C1-C6, and grouped them according to the syntactic patterns attributed to them in each model. This results in a three-level classification of our six conditions for the interpretive model and two levels for the generative approach.

The two variables were included as predictors in an ANCOVA model, with log-transformed frequency, trial order, previous trial RT, and previous segment (of the same trial) RT as controls/covariates.

To control for type 1 error rate, we used nested models in log-likelihood ratio tests in order to determine the contributions of individual variables, a standard method for dealing with type 1 error in multiple regression models.

### Double Priming Analysis: ANCOVA 2.0

As pointed out in Tooley and Traxler (2010), priming effects without lexical repetition in comprehension were reported in the context of double primed sentences in Thothathiri and Snedeker's (2008a) eye-tracking experiments with ambiguous double object and dative constructions. Thus, we designed a second ANCOVA model in order to test whether structural priming with unambiguous active sentences might be aided or affected in trials where two previous primes of the same category precede the target trial.

Including the same variables as in the previous ANCOVA 1.0 model in (31)-(32), we constructed a new ANCOVA model 2.0 by adding the two new variables in (33) and their interaction with the variables associated with their respective models ( $V_I$ ,  $V_G$ ) in (32). For each new variable, trials were coded as follows:

- (33) a. If the trial was preceded by TWO trials of the same condition (same as each other, not as the current trial, according to the interpretive theory), then the trial was coded as the condition of those 2 preceding trials (e.g., "Preceded by 2 Unergatives"). Otherwise, the trial was coded as "N/A."  
 b. If the trial was preceded by TWO trials of the same condition (same as each other, not as the current trial, according to the generative theory), then the trial was coded as the condition of those 2 preceding trials (e.g., "Preceded by 2 DP/Root"). Otherwise, the trial was coded as "N/A."

### Model-Tailoring Analysis: Linear Mixed Effects Regression Model

One of the potential limitations of our ANCOVAs quantitative analyses has to do with the fact that the dependent variable in the generative model had fewer levels than in the interpretive model

**TABLE 4** | Log-transformed mean reading time (St. Dev.): Condition × Previous Condition.

	PREVIOUS CONDITION						Average
	C1	C2	C3	C4	C5	C6	
Condition C1	6.119 (0.636)	6.098 (0.616)	6.096 (0.628)	6.09 (0.645)	6.093 (0.624)	6.084 (0.594)	6.097
C2	6.096 (0.588)	6.091 (0.626)	6.069 (0.573)	6.113 (0.602)	6.084 (0.622)	6.092 (0.649)	6.091
C3	6.054 (0.558)	6.036 (0.608)	6.057 (0.575)	6.054 (0.594)	6.052 (0.580)	6.048 (0.605)	6.050
C4	6.049 (0.609)	6.053 (0.574)	6.054 (0.613)	6.071 (0.572)	6.053 (0.637)	6.054 (0.537)	6.056
C5	6.074 (0.603)	6.066 (0.609)	6.064 (0.614)	6.068 (0.598)	6.093 (0.578)	6.086 (0.597)	6.075
C6	6.071 (0.608)	6.076 (0.590)	6.067 (0.612)	6.065 (0.593)	6.068 (0.575)	6.085 (0.578)	6.072
Average	6.077	6.070	6.068	6.077	6.074	6.075	

of Marantz, which could perhaps inherently restrict its ability to capture variance. To avoid this, we designed a linear mixed effects regression model that would test for syntactic priming on the basis of the grouping of conditions in each model. We took the same control variables as in our previous ANCOVA analyses, and coded two additional binary variables for each model, as in (34).

- (34) a.  $V_I$ -Binary Priming (coded as 1 for primed, 0 for unprimed)
- b.  $V_G$ -Binary Priming (same coding scheme)

Based on the predictions of each model in terms of priming relations between conditions as depicted in **Table 2**, we coded the variables in (34) as the two primary variables in (35).

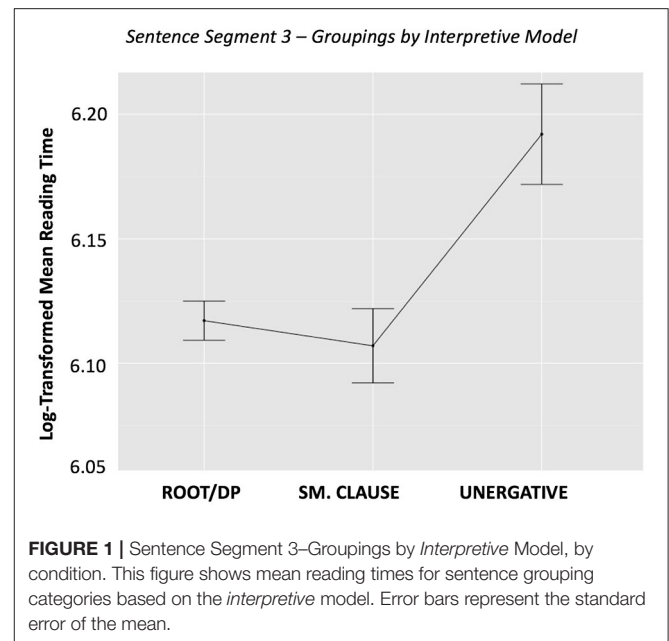
- (35) a. Primed: Anything that has a check mark in **Table 2** was coded as 1
- b. Unprimed: Anything that has a cross mark in **Table 2** was coded as 0

We coded two other binary variables for each model,  $V_{I(G)}$ - Same Previous and  $V_{I(G)}$  Same Two Previous [see (30a) and (30b) respectively]. Note that the subscript I(G) indicates that there were two corresponding variables calculated, one based on each model.

- (36) a.  $V_{I(G)}$ -Same Previous = binary variable coded 1 for trials where the previous trial was the same condition as the current, based on the respective model; 0 if not
- b.  $V_{I(G)}$ -Same Two Previous = binary variable coded 1 for trials where the previous two trials were the same condition as the current, based on the respective model; 0 if not

## RESULTS

**Table 4** shows the log-transformed mean response time and standard deviation for all individual conditions in all conditions of individual priming by condition (6 target × 6 previous).



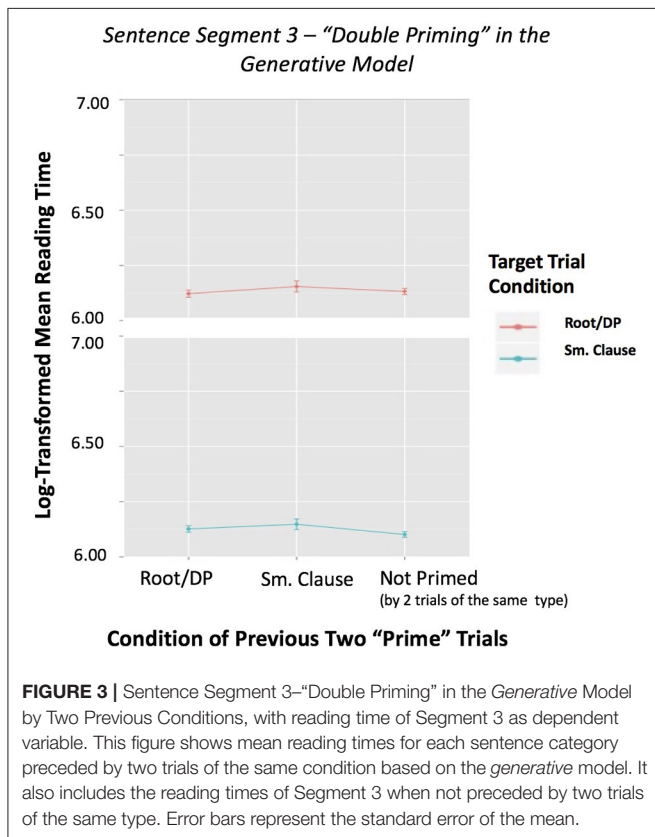
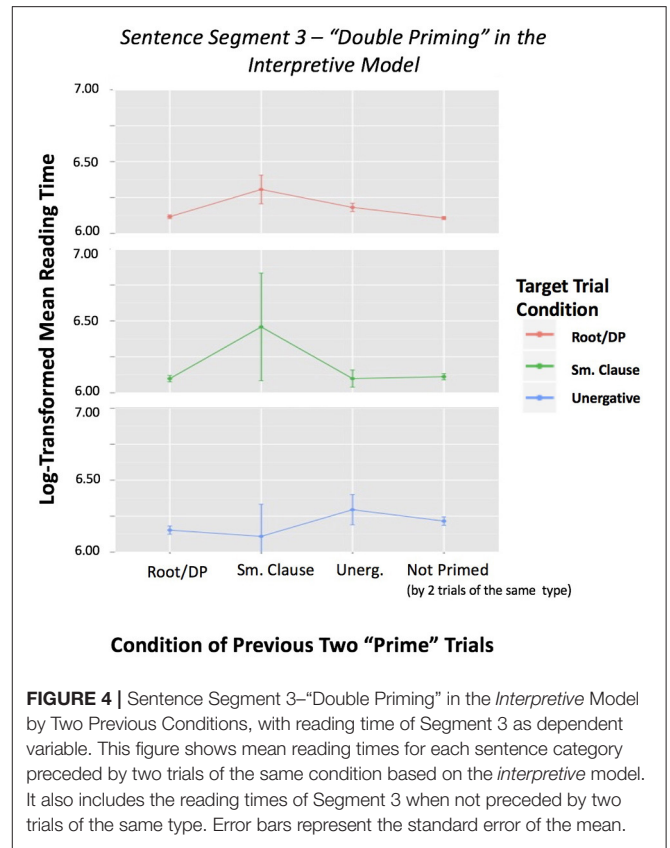
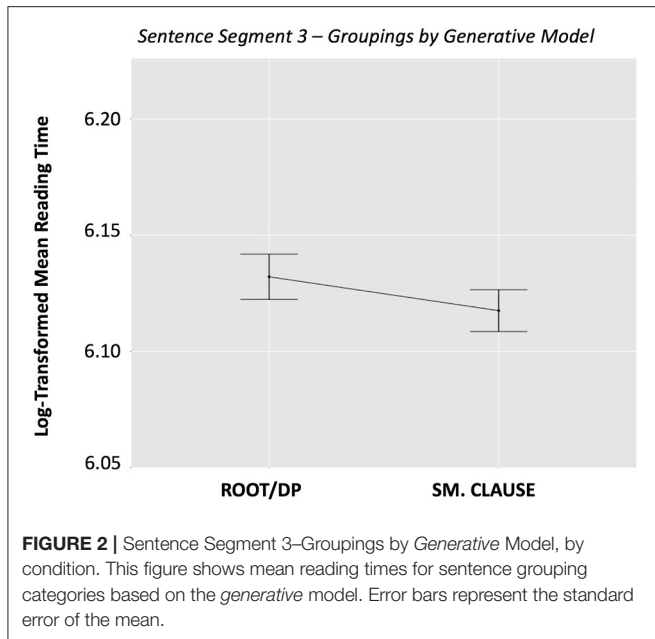
**FIGURE 1** | Sentence Segment 3-Groupings by *Interpretive Model*, by condition. This figure shows mean reading times for sentence grouping categories based on the *interpretive model*. Error bars represent the standard error of the mean.

The single priming analysis, ANCOVA 1.0, revealed that categorization based on the interpretive model ( $V_I$  - Sentence Type) was a significant predictor of reading time in Segment 3 ( $p = 0.012$ ). In contrast, categorization based on the generative model ( $V_G$ -Sentence Type) did not significantly affect reading time ( $p = 0.1379$ ). The raw reading times (i.e., not taking into account random effects or nuisance variables) are graphed in **Figures 1, 2**. The graphs should be interpreted cautiously, as they do not reflect the influence of random intercepts or nuisance variables, which were included in the ANCOVA, and thus are subject to potential confounds.

In the double priming analysis, ANCOVA 2.0, a “full” statistical model, i.e., one including the interaction between sentence type and previous (x2) sentence type, was tested against a model excluding the respective interaction terms, for both the generative and the interpretive theories. This initially gave us a null result. So, the contribution of the interpretive model interaction was not significant ( $p = 0.649$ ), nor was the contribution of the generative model interaction ( $p = 0.863$ ).

However, when we removed the random effects structure, keeping trial order as a covariate, we obtained again significant effects. The contribution of the *interpretive model* interaction was significant ( $p = 0.0037$ ), whereas the contribution of the *generative model* interaction was not significant ( $p = 0.756$ ). Even though these results should be interpreted with caution due to the simplified status of the model, they tentatively show a stronger predictive power of the *interpretive approach*. **Figures 3, 4** depict the interaction of sentence type by previous sentence type, according to each of the two models, with reading time of Segment 3 as the dependent variable.

As for our last statistical analysis, our model-tailoring analysis, no statistically significant effects were found in the linear mixed effects regression model, regardless of whether the random effects structure is included in the model, as shown in (37-38).



- (37) Without considering random effects
- Interpretive* model ( $p = 0.1078$ )
  - Generative* model ( $p = 0.2999$ )

- (38) With random effects
- Interpretive* model ( $p = 0.1766$ )
  - Generative* model ( $p = 0.565$ )

## DISCUSSION

Our first ANCOVA 1.0 analysis on single priming effects revealed a distinction between the two models. As shown in the relevant plots, there is no significant separation between conditions for the *generative* model, but we do observe separation in the *interpretive* model, particularly for the unergatives. In that sense, this effect of the *interpretive* model may be primarily driven by the fact that, in this approach, unergatives are considered to be their own category, whereas they are integrated in one of the groupings in the *generative* model, together with cognate object structures and verbs of creation.

Although the initial ANCOVA 2.0 analysis on double priming revealed no significant effects, after removing the random effects we observe a stronger predictive power of the *interpretive* approach. **Figure 3** shows no evidence that some set of V NP PP structures, those grouped under location/locatum sentences (C4) and strong transitives (C5), behaves like a small clause (SC) or that unergatives (C1) look like transitives (C2-C3) in the *generative* model. However, in **Figure 4**, we can observe effects for the small clause condition for the *interpretive* model. That is, two small clause sentences (C6) before a small clause sentence (C6) causes a slow-down in the reading times of Segment 3,

while two standard V NP PP sentences (C2-C3-C4-C5) before a small clause type sentence (C6) causes a significant speed up in Segment 3 reading times<sup>7</sup>.

Even though results were not significant, nor even trending, the linear mixed effects regression model is likely our most reliable model, because we have reduced the number of levels for the variables we are testing to just two for both models. It is worth noting that, as shown in (37) and (38), the effect size for the interpretive model is consistently larger than that of the generative approach, and the *p*-values of the interpretive model are consistently smaller, regardless of whether the random effects structure is included in the model.

We should note that even though the linear mixed effects regression model is most likely a more unbiased analysis, it does not allow us to investigate differences in priming between conditions, which is what we did in our ANCOVA 1.0, nor does it allow us to look at the interaction between the trial type and the prime type, as we showed in **Figures 3, 4**, resulting from our ANCOVA 2.0. Thus, the different statistical models we have employed do not exclude each other, but rather complement each other's limitations and they all seem to point toward a stronger predictive power of the *interpretive* approach.

## CONCLUSIONS

We have employed the experimental method to assess two competing linguistic accounts of the syntactic representation of the argument structure of (in)transitive structures on the basis of their divergent predictions with respect to sentence processing under conditions of syntactic priming. The design of our experiment makes use of on-line behavioral methods like self-paced reading, experimental techniques like priming, quantitative tools like frequency-based corpora, and sophisticated statistical control typical of experimental research in cognitive science to obtain reading time measures that allow us to effectively characterize theories about the representation and processing of syntactic phenomena. We have obtained significant results that point to a stronger predictive power of Marantz's interpretive theory over Acedo-Matellán and Mateu's generative model. Likewise, we have found no evidence in favor of the main claims of the generative analysis that some set of V NP PP structures behave like the small clauses or that unergatives are underlying transitives.

We have made a novel use of structural priming as a tool to discriminate among linguistic theories. A second novelty of the experiment lies in the structures we focus on, i.e., the empirical domain of the study. Whereas the central empirical issue in structural priming studies has mostly been how ambiguities arise and are resolved in incrementally disambiguating sentence processing, our empirical focus is on processing basic active simple (in)transitive structures. One of our main findings is thus

that there is structural priming in comprehension between basic structures without lexical boost.

To conclude, our controlled behavioral experimentation supports quantitative approaches to the study of I-language that advocate for the complementarity of psycholinguistic and theoretical methodologies to help us determine the nature of linguistic phenomena.

## LIMITATIONS AND FURTHER RESEARCH

As already mentioned above, one of the potential limitations of the model variables relates to the number of levels, three in Marantz's *interpretive* model vs. two in Acedo-Matellán and Mateu's *generative* model in the ANCOVAs, which could inherently restrict their ability to capture variance and as a consequence have an effect of grouping conditions by theory on our findings. Note, however, that the analysis where we test the ungrouped condition variable, i.e., the variable coded as 1–6, with six levels, is likewise not significant ( $p = 0.11$ ). Thus, it does not seem that adding more levels to the categorical predictor would improve the analysis. Yet, we should still interpret these results cautiously.

More data may be needed to see separation between the other conditions in the interpretive model or in the generative model, but that will likely be a focus of the future of this project. With respect to our ANCOVA 2.0, we only had few trials preceded by 2 trials of the same condition as the current trial, therefore, more data must be gathered to obtain reliable results in this direction. At this point, while we have preliminary effects showing the interpretive model is a better predictor, this appears to be only based on one aspect of the model, and we may not currently have enough statistical power to look at all aspects of the model.

We have also detected an unexpected slow-down in response times for primed trials that must be further investigated.

## AUTHOR CONTRIBUTIONS

IO and AM designed the research; IO drafted the work; all authors critically revised the drafts; VS and AM designed the MTurk; VS ran the MTurk and the statistical models on R; VS pre-processed the data. All authors analyzed the data, interpreted the results and wrote the final paper.

## FUNDING

This work has been supported by NYU Abu Dhabi Research Institute Grant G1001, the Spanish Ministry of Economy and Competitiveness (FFI2016-80142-P), the Spanish Ministry of Education ("José Castillejo" mobility grant CAS14/00418) and grant SGR1546-2014 (AGAUR).

## ACKNOWLEDGMENTS

We are grateful to two reviewers and to the audiences of the 38th Annual Meeting of the *Deutsche Gesellschaft*

<sup>7</sup>It should be noted that our graphs do not technically include random effects and order, nor any of our other control variables; to the best of our knowledge, there is no way to cleanly integrate them into a plot. This is why the graphs may not look particularly informative.

für *Sprachwissenschaft* DGfS in Konstanz (Germany), and the Seminar of the CLT-UAB in Barcelona in 2016 for helpful comments and suggestions. Thanks to L. Stockall and the members of the NeLLab NYU/NYUAD.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01311/full#supplementary-material>

## REFERENCES

- Acedo-Matellán, V. (2010). *Argument Structure and the Syntax-Morphology Interface. A Case Study in Latin and Other Languages*. Dissertation, Barcelona: Universitat de Barcelona.
- Acedo-Matellán, V. (2014). "Can we dance without doing a dance?" in *Structuring the Argument: Multidisciplinary Research on Verb Argument Structure*, eds A. I. Bachrach, Roy, and L. Stockall (Amsterdam: John Benjamins), 23–43.
- Acedo-Matellán, V., and Mateu, J. (2013). Satellite-framed Latin vs. verb-framed Romance: a syntactic approach. *Probus* 25, 227–265. doi: 10.1515/probus-2013-0008
- Arai, M., Van Gompel, R. P., and Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cogn. Psychol.* 54, 218–250. doi: 10.1016/j.cogpsych.2006.07.001
- Bock, K. (1986). Syntactic persistence in language production. *Cogn. Psychol.* 18, 355–387. doi: 10.1016/0010-0285(86)90004-6
- Bock, K., Dell, G. S., Chang, F., and Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition* 104, 437–458. doi: 10.1016/j.cognition.2006.07.003
- Bock, K., and Griffin, Z. M. (2000). The persistence of structural priming: transient activation or implicit learning? *J. Exp. Psychol.* 129, 177–192. doi: 10.1037/0096-3445.129.2.177
- Bock, K., and Loebell, H. (1990). Framing sentences. *Cognition* 5, 1–39. doi: 10.1016/0010-0277(90)90035-1
- Bock, K., Loebell, H., and Morey, R. (1992). From conceptual roles to structural relations: bridging the syntactic cleft. *Psychol. Rev.* 99:150. doi: 10.1037/0033-295X.99.1.150
- Branigan, H. P., Pickering, M. J., Stewart, A. J., and McLean, J. F. (2000). Syntactic priming in spoken production: linguistic and temporal interference. *Mem. Cogn.* 28, 1297–302. doi: 10.3758/BF03211830
- Branigan, H. P., Pickering, M. J., Liversedge, S. P., Stewart, A. J., and Urbach, T. P. (1995). Syntactic priming: investigating the mental representation of language. *J. Psycholinguist. Res.* 24, 489–506. doi: 10.1007/BF02143163
- Branigan, H. P., Pickering, M. J., and McLean, J. F. (2005). Priming prepositional-phrase attachment during comprehension. *J. Exp. Psychol.* 31, 468–481. doi: 10.1037/0278-7393.31.3.468
- Carminati, M. N., van Gompel, R. P., Scheepers, C., and Arai, M. (2008). Syntactic priming in comprehension: the role of argument order and animacy. *J. Exp. Psychol.* 34, 1098–1110. doi: 10.1037/a0012795
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Westport, CT: Praeger.
- Cinque, G., and Rizzi, L. (2010). *Mapping Spatial PPs: The Cartography of Syntactic Structures*. Oxford: Oxford University Press.
- Clark, E. V., and Clark, H. H. (1979). When nouns surface as verbs. *Language* 55, 767–811. doi: 10.2307/412745
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 520 Million Words, 1990–Present*. Available online at: <http://corpus.byu.edu/coca/>
- Ferreira, V. S., and Bock, K. (2006). The functions of structural priming. *Lang. Cogn. Process.* 21, 1011–1029. doi: 10.1080/01690960600824609
- Hale, K., and Keyser, S. J. (1993). *On Argument Structure and the Lexical Expression of Syntactic Relations. The View from Building*. Cambridge, MA: The MIT Press.
- Hale, K., and Keyser, S. J. (1998). "The basic elements of argument structure," in *MIT Working Papers in Linguistics* (Cambridge, MA), 32, 73–118.
- Hale, K., and Keyser, S. J. (2002). *Prolegomenon to a Theory of Argument Structure*. Cambridge, MA: The MIT Press.
- Hare, M. L., and Goldberg, A. E. (1999). "Structural priming: purely syntactic?" in *Proceedings of the 21st Annual Meeting of the Cognitive Science Society*, eds M. Hahn and S. C. Stones (Mahwah, NJ: Erlbaum), 208–211.
- Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., and Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: evidence from written and spoken dialogue. *J. Mem. Lang.* 58, 214–238. doi: 10.1016/j.jml.2007.07.003
- Horn, L. R. (1980). Affixation and the unaccusative hypothesis. *Chic. Linguist. Soc.* 16, 134–146.
- Jackendoff, R. (1973). "The base rules for prepositional phrases," in *A Festschrift for Morris Halle*, eds S. Anderson and P. Kiparsky (New York, NY: Holt, Rinehart and Winston), 345–366.
- Katz, J. (1971). Generative semantics is interpretive semantics. *Linguist. Inq.* 2, 313–331.
- Korhonen, A., Krymowski, Y., and Briscoe, T. (2006). "A large subcategorization lexicon for natural language processing applications," in *Proceedings of LREC*, Vol. 6. (Genoa), 1015–1020.
- Loebell, H., and Bock, K. (2003). Structural priming across languages. *Linguistics* 41, 791–824. doi: 10.1515/ling.2003.026
- Mahowald, K., James, A., Futrell, R., and Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *J. Mem. Lang.* 91, 5–27. doi: 10.1016/j.jml.2016.03.009
- Marantz, A. (2005). Generative linguistics within the cognitive neuroscience of language. *Linguist. Rev.* 22, 429–445. doi: 10.1515/tlir.2005.22.2-4.429
- Marantz, A. (2011). "Syntactic approaches to argument structure without incorporation," *Talk Presented at the Workshop Structuring the argument, Structures Formelles du Langage UMR 7023 Paris 8/CNRS*, Paris, September 5–7.
- Marantz, A. (2013). Verbal argument structure: events and participants. *Lingua* 130, 152–168. doi: 10.1016/j.lingua.2012.10.012
- Mateu, J. (2002). *Argument Structure. Relational Construal at the Syntax-Semantics Interface*. Dissertation, Bellaterra: Universitat Autònoma de Barcelona.
- Mateu, J., and Acedo-Matellán, V. (2012). "The manner/result complementarity revisited: a syntactic approach," in *The End of Argument Structure*, eds M. C. Cuervo and Y. Roberge (Brill Online), 209–228.
- Melinger, A., and Dobel, C. (2005). Lexically-driven syntactic priming. *Cognition* 98, B11–B20. doi: 10.1016/j.cognition.2005.02.001
- Pickering, M. J., and Branigan, H. P. (1998). The representation of verbs: evidence from syntactic priming in language production. *J. Mem. Lang.* 39, 633–651. doi: 10.1006/jmla.1998.2592
- Pickering, M. J., and Branigan, H. P. (1999). Syntactic priming in language production. *Trends Cogn. Sci.* 3, 136–141. doi: 10.1016/S1364-6613(99)01293-0
- Pickering, M. J., Branigan, H. P., and McLean, J. F. (2002). Constituent structure is formulated in one stage. *J. Mem. Lang.* 46, 586–605. doi: 10.1006/jmla.2001.2824
- Pickering, M. J., and Ferreira, V. S. (2008). Structural priming: a critical review. *Psychol. Bull.* 134, 427–459. doi: 10.1037/0033-2909.134.3.427
- Pickering, M. J., McLean, J. F., and Branigan, H. P. (2013). Persistent structural priming and frequency effects during comprehension. *J. Exp. Psychol.* 39, 890–897. doi: 10.1037/a0029181
- Pickering, M. J., and Traxler, M. J. (2004). "Grammatical repetition and garden-path effects," in *Talk presented at the 17th annual CUNY Conference on Human Sentence Processing* (Maryland, MD).
- Poeppl, D., and Embick, D. (2005). "The relation between linguistics and neuroscience," in *Twenty-first Century Psycholinguistics: Four Cornerstones*, ed A. Cutler (Hillsdale, NJ: Lawrence Erlbaum Associates), 103–120.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. Vienna. Available online at: URL <http://www.R-project.org>.

- Segaert, K., Kempen, G., Petersson, K. M., and Hagoort, P. (2013). Syntactic priming and the lexical boost effect during sentence production and sentence comprehension: an fMRI study. *Brain Lang.* 124, 174–183. doi: 10.1016/j.bandl.2012.12.003
- Segaert, K., Menenti, L., Weber, K., Petersson, K. M., and Hagoort, P. (2012). Shared syntax in language production and language comprehension—an fMRI study. *Cereb. Cortex* 24, 1662–1670. doi: 10.1093/cercor/bhr249
- Talmy, L. (1975). “Figure and ground in complex sentences,” in *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society, Vol. 1, No. 1* (Berkeley, CA: Linguistics Society).
- Thothathiri, M., and Snedeker, J. (2006). “Children’s representation of verbs: evidence from priming during online sentence comprehension,” in *Proceedings of the 30th Annual Boston University Conference on Language Development*, eds D. Bamman, M. Tatiana, and C. Zaller (Somerville, MA: Cascadilla), 643–652.
- Thothathiri, M., and Snedeker, J. (2008a). Syntactic priming during language comprehension in three- and four-year-old children. *J. Mem. Lang.* 58, 188–213. doi: 10.1016/j.jml.2007.06.012
- Thothathiri, M., and Snedeker, J. (2008b). Give and take: syntactic priming during spoken language comprehension. *Cognition* 108, 51–68. doi: 10.1016/j.cognition.2007.12.012
- Tooley, K. M., and Bock, K. (2014). On the parity of structural persistence in language production and comprehension. *Cognition* 132, 101–136. doi: 10.1016/j.cognition.2014.04.002
- Tooley, K. M., Swaab, T. Y., Boudewyn, M. A., Zirnstein, M., and Traxler, M. J. (2014). Evidence for priming across intervening sentences during on-line sentence comprehension. *Lang. Cogn. Process.* 29, 289–311. doi: 10.1080/01690965.2013.770892
- Tooley, K. M., and Traxler, M. J. (2010). Syntactic priming effects in comprehension: a critical review. *Lang. Linguist. Compass* 4, 925–937. doi: 10.1111/j.1749-818X.2010.00249.x
- Tooley, K. M., Traxler, M. J., and Swaab, T. Y. (2009). Electrophysiological and behavioral evidence of syntactic priming in sentence comprehension. *J. Exp. Psychol.* 35, 19–45. doi: 10.1037/a0013984
- Traxler, M. J. (2008a). Structural priming among prepositional phrases: evidence from eye movements. *Mem. Cogn.* 36, 659–674. doi: 10.3758/MC.36.3.659
- Traxler, M. J. (2008b). Lexically independent priming in online sentence comprehension. *Psychon. Bull. Rev.* 15, 149–155. doi: 10.3758/PBR.15.1.149
- Traxler, M. J., and Tooley, K. M. (2007). Lexical mediation and context effects in sentence processing. *Brain Res.* 1146, 59–74. doi: 10.1016/j.brainres.2006.10.010
- Traxler, M. J., and Tooley, K. M. (2008). Priming in sentence comprehension: strategic or syntactic? *Lang. Cogn. Process.* 23, 609–645. doi: 10.1080/01690960701639898
- Traxler, M. J., Tooley, K. M., and Pickering, M. J. (2014). Syntactic priming during sentence comprehension: evidence for the lexical boost. *J. Exp. Psychol.* 40, 905–918. doi: 10.1037/a0036377
- Trueswell, J. C., and Kim, A. E. (1998). How to prune a garden path by nipping it in the bud: fast priming of verb argument structure. *J. Mem. Lang.* 39, 102–123. doi: 10.1006/jmla.1998.2565
- Warren, T., and Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition* 85, 79–112. doi: 10.1016/S0010-0277(02)00087-2
- Wechsler, S. (1989). Accomplishments and the prefix re-. *NELS* 19, 419–434.
- Wittenberg, E., Jackendoff, R., Kuperberg, G., Paczynski, M., Snedeker, J., and Wiese, H. (2014). “The mental representation and processing of light verbs,” in *Structuring the Argument: Multidisciplinary Research on Verb Argument Structure*, eds A. Bachrach, I. Roy, and L. Stockall (Amsterdam: John Benjamins), 61–80.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Oltra-Massuet, Sharpe, Neophytou and Marantz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# On the Nature of Clitics and Their Sensitivity to Number Attraction Effects

Mikel Santesteban\*, Adam Zawiszewski, Kepa Erdocia and Itziar Laka

Department of Linguistics and Basque Studies, University of the Basque Country (UPV/EHU), Vitoria-Gasteiz, Spain

Pronominal dependencies have been shown to be more resilient to attraction effects than subject-verb agreement. We use this phenomenon to investigate whether antecedent-clitic dependencies in Spanish are computed like agreement or like pronominal dependencies. In Experiment 1, an acceptability judgment self-paced reading task was used. Accuracy data yielded reliable attraction effects in both grammatical and ungrammatical sentences, only in singular (but not plural) clitics. Reading times did not show reliable attraction effects. In Experiment 2, we measured electrophysiological responses to violations, which elicited a biphasic frontal negativity-P600 pattern. Number attraction modulated the frontal negativity but not the amplitude of the P600 component. This differs from ERP findings on subject-verb agreement, since when the baseline matching condition obtained a biphasic pattern, attraction effects only modulated the P600, not the preceding negativity. We argue that these findings support cue-retrieval accounts of dependency resolution and further suggest that the sensitivity to attraction effects shown by clitics resembles more the computation of pronominal dependencies than that of agreement.

**Keywords:** clitics, agreement, pronouns, object agreement, attraction effects, sentence processing, cue-based retrieval

## OPEN ACCESS

### Edited by:

Aritz Irurtzun,  
Centre National de la Recherche  
Scientifique (CNRS), France

### Reviewed by:

Brian Dillon,  
University of Massachusetts Amherst,  
United States  
Matthew Wagers,  
University of California, Santa Cruz,  
United States

### \*Correspondence:

Mikel Santesteban  
mikel.santesteban@ehu.eus;  
msantesteban@gmail.com

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 23 January 2017

**Accepted:** 15 August 2017

**Published:** 05 September 2017

### Citation:

Santesteban M, Zawiszewski A,  
Erdocia K and Laka I (2017) On  
the Nature of Clitics and Their  
Sensitivity to Number Attraction  
Effects. *Front. Psychol.* 8:1470.  
doi: 10.3389/fpsyg.2017.01470

## INTRODUCTION

Discovering the dependency relations between different elements of a sentence allows us to disentangle its meaning. In these dependency relations, verbal or nominal constituents match in certain features (i.e., number, person and/or gender) with another constituent of the sentence (Corbett, 2006). One of the most frequently studied dependency is that between a subject and a verb, where the features of the subject (e.g., the number) determine the form of the verb (e.g., *the key is...* vs. *the keys are...*) (see Bock and Middleton, 2011 for a review). In this paper we investigate a type of syntactic dependency that has received little attention in psycholinguistics: antecedent-clitic relations. There is debate in linguistics regarding the nature of clitics, where clitics are argued to be either pronouns or agreement morphemes. Our main objective is to experimentally explore the nature of antecedent-clitic dependencies. For that purpose, we use *agreement attraction*, a phenomenon showing that the presence of alternative candidates can disrupt the computation of dependency relations between two elements (Bock and Miller, 1991; Nicol et al., 1997). More specifically, we explore whether antecedent-clitic dependencies show similar behavioral (Experiment 1) and electrophysiological (Experiment 2) patterns of number

agreement attraction as those previously reported for subject-verb agreement relations or as those reported for antecedent-pronoun relations.

## Why Antecedent-Clitic Dependencies?

The nature of Romance clitics has been much debated in Generative Linguistics since the seminal works by Kayne (1975) and Zwicky (1977), but experimental evidence regarding how they are processed is scarce. The status of clitics and particularly Romance clitics are an important subject of research in generative linguistics due to their intermediate/mixed behavior between independent pronouns and affixed agreement morphemes. In the case of the Spanish object-clitics we studied, they agree with their antecedent in number [*Anna vió la novela<sub>fem.sg</sub>/las novelas<sub>fem.pl</sub> y la<sub>fem.sg</sub>/las<sub>fem.pl</sub> compró*; “Anna saw the novel<sub>fem.sg</sub>/s<sub>fem.pl</sub> and (she) bought it<sub>fem.sg</sub>/them<sub>fem.pl</sub>”], and gender [*Anna vió el libro<sub>masc.sg</sub>/los libros<sub>masc.pl</sub> y lo<sub>masc.sg</sub>/los<sub>masc.pl</sub> compró*; “Anna saw the book/s and (she) bought it<sub>masc.sg</sub>/them<sub>masc.pl</sub>”], unlike verbal inflection that agrees in person and number. These object-clitics correspond to the object arguments of the sentences’ main verb *comprar* (“to buy”). Hence, like pronouns, Spanish object clitics agree in gender and not person, satisfy verbal subcategorization properties and behave as arguments of the verb. However, like agreement (inflectional) morphemes, clitics are unstressed and affixed to the verb. In generative linguistics, there are two main competing approaches accounting for the nature of clitics:

Kayne (1975) originally proposed that clitics were syntactically independent elements in what we will refer to as the *Clitics as Pronouns Hypothesis*: clitics are pronoun noun phrases (NPs) generated at argument position that attach to the verb in the course of the derivation. In this view, NP-clitic dependencies are a case of referential co-dependency and the clitic occupies the argument position (Torrego, 1988; Uriagereka, 1995; Sportiche, 1998; Anagnostopoulou, 2003; Marchis and Alexiadou, 2013, among others). In a variant of this hypothesis, the clitic is generated in its surface position, while the argument position is filled by the empty pronominal *pro* (Strozer, 1976; Rivas, 1977; Jaeggli, 1982; Borer, 1984; among others). On the other hand, according to what we will refer to as the *Clitics as Agreement Hypothesis*, pronominal clitics are agreement morphemes, part of Inflection and not generated in argument position (e.g., Jaeggli, 1986; Suñer, 1988; Fernández Soriano, 1989; Monachesi, 2005 among others).

Our main objective is to contribute to better understanding the nature of clitics by testing whether and to what extent the behavioral and electrophysiological pattern found during clitic processing resembles that reported previously in the literature for verb agreement, or whether it aligns better with the processing patterns of pronominal concord. To that end, we explore (i) whether, in behavioral measures, antecedent-clitic dependencies are prone to number attraction effects similar to those found in subject-verb agreement, or whether they are more resilient to these effects as antecedent-pronoun dependencies are (see further discussion about this issue in next section); and (ii) whether, in electrophysiological measures, they elicit the same electrophysiological indexes of attraction as those previously reported for subject-verb agreement.

## On Number Attraction Effects

The study of the contexts where attraction phenomena occur during language production has shed light on the main factors involved in agreement processing: in sentence preambles such as *The key to the cabinet(s)...*, speakers produce more number agreement errors completing preambles containing an *attractor* noun that does not match (i.e., *cabinets*) in number with the agreement controller (i.e., the head noun *key*), than when the attractor matches (Bock and Miller, 1991; see Bock and Middleton, 2011; Franck, 2011 for exhaustive reviews of attraction effects in various types of agreement dependencies). Research on attraction effects in language comprehension is much more scarce than in production, and it has considered almost exclusively subject-verb agreement (in English: Nicol et al., 1997; Pearlmutter et al., 1999; Pearlmutter, 2000; Wagers et al., 2009; Shen et al., 2013; in Dutch: Kaan, 2002; Chen et al., 2007; Severens et al., 2008; in Spanish: Acuña-Fariña et al., 2014; Lago et al., 2015; in French: Franck et al., 2015). However, recent studies have also explored antecedent-reflexive pronoun concord (Dillon et al., 2013, 2014, 2016; Jäger et al., 2015; Patil et al., 2016; Parker and Phillips, 2017; for a thorough literature review on attraction effects in subject-verb and antecedent-pronoun dependencies, see Jäger et al., 2017).

Early studies adopted the *feature percolation hypothesis* postulated to account for attraction effects in language production. According to this account, attraction effects in both production and comprehension occur because the number features of the attractor noun can erroneously percolate over the number features of the agreement controller, which results in an erroneous number representation of the agreement controller (e.g., Nicol et al., 1997; Pearlmutter et al., 1999; Pearlmutter, 2000).

More recently, it has been proposed that attraction effects are best accounted for by means of a similarity-based interference model (Badecker and Kuminiak, 2007; Wagers et al., 2009; see Dillon et al., 2013; Jäger et al., 2017; for computational simulations of the model) inspired in the ACT-R model (Lewis and Vasishth, 2005). According to this model, dependency relations are established by retrieving from memory the agreeing dependents. When the agreeing element (e.g., a verb or a clitic) is encoded, it engages a cue-based retrieval mechanism to search for a matching controller in memory. But this retrieval mechanism is susceptible to similarity-based interference from other items in memory. Hence, when a distracting element that carries similar features (e.g., semantic, structural features) as the controller is present in the sentence, interference occurs because the distracting element might be misidentified as the controller. Importantly, this model predicts attraction effects to be only present or to be larger during the processing of ungrammatical than grammatical sentences. Wagers et al. (2009) suggested two options for cue-retrieval mechanisms to account for these asymmetric effects: (a) encountering the agreeing element engages retrieval mechanisms that retrieve number-matching NPs but (almost) never retrieve partially matching ones (i.e., a number mismatching attractor in grammatical sentences); or (b) the correct agreeing element form is predicted after encountering the controller NP and the cue-based reanalysis

process ensues almost exclusively when ungrammaticality is detected.

However, several studies report the presence of number attraction effects in both grammatical and ungrammatical sentences. In these studies, sentence acceptability, self-paced reading for comprehension and eye-tracking measures showed that participants are slower reading or accepting grammatical sentences with a singular subject and a plural attractor (e.g., *The author of the speeches was. . . vs. The author of the speech was. . .*) than accepting sentences where both NPs were singular (Nicol et al., 1997; Pearlmutter et al., 1999; Pearlmutter, 2000; Acuña-Fariña et al., 2014). In contrast, for ungrammatical sentences, mismatching attractors have been shown to elicit faster reading times as compared to matching ones in self-paced reading tasks (Pearlmutter et al., 1999; Wagers et al., 2009; Franck et al., 2015; Lago et al., 2015) and eye-tracking measures (Dillon et al., 2013). That is, attraction effects interfere in the processing of the agreement controller in grammatical sentences but facilitate it in ungrammatical ones. However, Wagers et al. (2009) identified a confound variable that might have led to the interference attraction effects reported in grammatical sentences: since in all these studies attractors and agreeing verbs were adjacent, the interference effects observed at the verb might be due to carry-over effects of the slower times needed to process the morphologically marked plural rather than unmarked singular attractors.

Nevertheless, in a recent study, Franck et al. (2015) showed both facilitation and interference attraction effects in grammatical sentences where the attractor and the verb were not adjacent and they suggested that experimental design factors might affect the direction of the effect. In a self-paced reading for comprehension task in French only including grammatical sentences (Experiment 1), they reported attraction facilitation effects. In contrast, in a speeded acceptability judgment task, participants showed attraction interference effects (slower acceptability judgments) when judging both grammatical and ungrammatical sentences containing number mismatching attractors, as compared to matching ones. Franck et al. (2015) interpreted their results as evidence that different behavioral tasks tap into different processes: while self-paced reading taps structure building processes, grammaticality judgment taps into later processes of agreement computation. Either way, the fact that attraction effects were detected in both grammatical and ungrammatical sentences might support feature percolation accounts.

However, many recent studies found reliable attraction effects in ungrammatical sentences but not in grammatical ones, favoring similarity-based interference accounts (Wagers et al., 2009; Dillon et al., 2013; Lago et al., 2015). This *grammatical vs. ungrammatical asymmetry* of attraction effects was interpreted as the main evidence that attraction effects are mainly due to similarity-based interference effects during the retrieval of the cues necessary to build dependency relations (e.g., Lewis and Vasishth, 2005), and not due to a faulty representation of the agreement controller, as suggested by the feature percolation account. As reviewed in the next section, electrophysiological evidence of attraction effects replicated the

grammatical asymmetry of attraction effects (e.g., Kaan, 2002; Shen et al., 2013; Tanner et al., 2014, 2016).

Morphological markedness plays a crucial role during agreement attraction in comprehension: attraction effects are either only found in singular, but not plural agreement (Nicol et al., 1997; Wagers et al., 2009, in acceptability and self-paced reading data), or are larger in singular than plural agreement (Acuña-Fariña et al., 2014, in eye-tracking measures). These findings replicate the number markedness effects also reported in production studies (Bock and Miller, 1991; Bock and Cutting, 1992; Bock and Eberhard, 1993; Eberhard, 1997), suggesting that morphologically marked plural distractors are stronger attractors than non-marked singular ones in both modalities. Thus, attraction effects might sometimes be obscured and delayed due to carry-over effects of plural attractors when the attractor and the agreeing element are adjacent. However, those carry-over effects do not last long: they can be avoided by including a word between the attractor and the verb (e.g., Wagers et al., 2009) and even when the attractor and the verb are adjacent, attraction effects are detected at the region following the verb (Pearlmutter et al., 1999).

All research reviewed above studied subject-verb agreement dependencies. But do attraction effects also affect the processing, and more particularly the comprehension of antecedent-pronoun dependencies? In production, pronoun-antecedent agreement seems to be as sensitive to attraction effects as subject-verb agreement is, but the former is more sensitive to notional number factors (e.g., Bock et al., 1999, 2004), suggesting that pronominal dependencies may rely more on the retrieval of the semantic/lexical representation of the antecedent. In comprehension, early studies exploring the role of grammatical constraints in antecedent-reflexive pronoun gender agreement showed that they are resilient to interference from other possible antecedent candidates (Nicol and Swinney, 1989; Sturt, 2003; *inter alia*). More recently, these findings have been replicated in studies that compared the magnitude of attraction effects in antecedent-reflexive pronoun vs. subject-verb agreement dependencies. In a reading for comprehension eye-tracking experiment, Dillon et al. (2013) showed reliable attraction effects for subject-verb agreement (shorter total reading times and fewer regressions to the critical agreement region were obtained in sentences containing mismatching attractors as compared to sentences containing matching ones, but only in ungrammatical sentences, replicating the grammatical asymmetry of attraction). No signs of attraction effects were found for reflexive pronouns (e.g., *The new executive who oversaw the middle manager/s apparently doubted himself/\*themselves. . .*). Dillon et al. (2013) interpreted the resilience of reflexive pronouns to attraction effects as evidence that subject-verb vs. antecedent-reflexive pronoun dependencies involve qualitatively different processes (see also Phillips et al., 2011). According to the authors, these different linguistic dependencies are sensitive to different linguistic features: (a) verbal agreement is a formal morphosyntactic mechanism to index the arguments of the verb, and feature retrieval is mainly driven by ranked morphological and structural cues (i.e., number feature and subjecthood cues, respectively); and (b) pronominal concord is a dependency

between two NPs and therefore antecedent retrieval is driven by syntactic (structural) cues.

Interestingly, recent eye-tracking studies show that although (English) reflexives are more resilient to attraction, they are indeed susceptible to it. For instance, Patil et al. (2016) showed that when the role of structural-cues such as subjecthood is controlled (e.g., both the antecedent of the reflexive and the attractor were subjects), attraction effects occurred when the attractor mismatched in morphological cues such as gender. Parker and Phillips (2017) also showed that no attraction effects occurred when the attractor mismatched in a single feature (i.e., gender) with the antecedent, but they did when the attractor mismatched in two features (e.g., gender and animacy, number and animacy or number and gender). These authors suggested that both subject-verb and antecedent-reflexive pronoun agreement engage similar cue-based retrieval mechanisms. However, following Dillon et al. (2013), Parker and Phillips (2017) suggested that reflexive pronoun dependencies weight structural cues more strongly than morphological cues, which precludes the erroneous retrieval of non-licensed antecedent candidates (see also Dillon et al., 2014, 2016).

In sum, behavioral measures show that subject-verb agreement comprehension is prone to attraction effects (Nicol et al., 1997; Pearlmutter et al., 1999; Pearlmutter, 2000; Wagers et al., 2009; Acuña-Fariña et al., 2014; Franck et al., 2015; Lago et al., 2015), but antecedent-pronoun dependencies are more resilient to these effects (Dillon et al., 2013; Patil et al., 2016; Parker and Phillips, 2017; see also Jäger et al., 2017 for a thorough review and discussion). Although attraction effects have been also reported in grammatical sentences, they are stronger and more consistent in ungrammatical sentences (e.g., Pearlmutter et al., 1999; Wagers et al., 2009; Lago et al., 2015), supporting similarity-based accounts. Next, we review the main findings of ERP studies on agreement and number attraction effects.

## ERP Correlates of Syntactic Dependency Processing

In general, when processing syntactic violations in subject-verb, object-verb, or antecedent-pronoun dependencies, three types of electrophysiological correlates have been reported in the ERP literature: Left Anterior Negativity (LAN), N400 and a centro-parietal positivity (P600) (for a detailed description and interpretation of each component see i.e., Bornkessel and Schlesewsky, 2006).

Most studies observed biphasic patterns with negative components (LAN/N400) followed by a positive component (P600). Some studies reported a biphasic LAN – P600 pattern for subject-verb agreement violations (Kutas and Hillyard, 1983; Osterhout and Mobley, 1995; De Vincenzi et al., 2003; Silva-Pereyra and Carreiras, 2007, among others) as well as for determiner-noun or noun-adjective gender agreement violations (Gunter et al., 2000; Deutsch and Bentin, 2001; Barber and Carreiras, 2005; Martin-Loeches et al., 2006; Molinaro et al., 2008). Other studies reported a biphasic N400-P600 pattern for subject-verb and object-verb agreement violations (Coulson et al., 1998; Zawiszewski and Friederici, 2009; Díaz et al., 2011; Zawiszewski et al., 2011) as well as for antecedent-pronoun

violations (Schmitt et al., 2002; Hammer et al., 2005, 2008; Lamers et al., 2006). Finally, some studies have also reported an isolated P600 component for subject-verb agreement violations (Osterhout et al., 1996; Nevins et al., 2007; Frenck-Mestre et al., 2008), for determiner-noun or noun-adjective gender agreement relations (Osterhout and Mobley, 1995; Osterhout et al., 1997; Foucart and Frenck-Mestre, 2011, 2012) and for antecedent-pronoun violations (Lamers et al., 2006, 2008; Silva-Pereyra et al., 2012; Xu et al., 2013; Rossi et al., 2014). As far as we know, no study has shown an isolated early negativity (N400 or LAN).

## ERP Correlates of Attraction Effects

Regarding the electrophysiological responses underlying number attraction effects, the available evidence is rather scarce and focused on subject-verb number agreement. To our knowledge, no study explored attraction effects in antecedent-clitic dependencies.

Electrophysiological indexes of attraction effects in subject-verb agreement are heterogeneous, but two main results have been observed: (a) electrophysiological indexes of agreement violation detection are less salient and harder to detect in sentences containing number mismatching attractors than matching attractors (Kaan, 2002; Chen et al., 2007; Severens et al., 2008; Shen et al., 2013; Tanner et al., 2014, 2016); and (b) the four studies that checked for asymmetrical attraction effects found an asymmetry: number mismatching attractors elicit a reduction of ERP components as compared to number matching ones in ungrammatical sentences, but not in grammatical ones (Kaan, 2002; Shen et al., 2013; Tanner et al., 2014, 2016).

Focusing on the studies that reported asymmetrical attraction effects, and thus support the cue-based retrieval account of agreement computation (Lewis and Vasishth, 2005; Wagers et al., 2009), Kaan (2002) investigated the effects of distance and number interference in subject agreement processing: Dutch participants performed an acceptability rating task in sentences containing subject and object NPs that either matched or mismatched in number. ERP responses following the critical verb revealed main grammaticality effects reflected by a bilateral negativity over central and posterior sites between 300 and 500 ms, and a P600 effect between 500–700 and 700–900 ms. A main number attraction effect was revealed by a significantly larger P600 component between 500 and 700 ms following subject agreement violations in sentences with only singular NPs (i.e., the control singular number matching condition) than in any other condition. Number mismatching attractors elicited a smaller P600 in singular subject agreement, but not in plural, replicating the number markedness effects (Eberhard, 1997; Nicol et al., 1997; Wagers et al., 2009). Finally, the modulation of the P600 related to attraction effects was asymmetrical, as it only occurred in ungrammatical sentences.

Tanner et al. (2014) provide behavioral and ERP evidence supporting the asymmetric pattern of attraction effects: English speaking participants showed a main P600 component elicited by subject-verb agreement violations, with attraction effects revealing a smaller P600 in sentences containing number mismatching attractors than number matching ones. These

attraction effects were asymmetrical: Participants showed a reliable P600 effect and were less accurate judging ungrammatical sentences that contained number mismatching rather than number matching attractor NPs. In contrast, they showed no P600 effect and were similarly accurate while judging the acceptability of grammatical sentences containing number matching and mismatching attractor NPs. These results obtained both with and without an adverb intervening between the attractor noun and the auxiliary verb (*The chemist with the test tube(s) (probably) is/\*are. . .*), suggesting that ERP indexes of attraction are resilient to carry-over effects of the plural attractor. In a recent study, Tanner et al. (2016) replicated this pattern of attraction effects revealing that number mismatching attractors reduce the magnitude of the P600 as compared to number matching ones in ungrammatical sentences.

Shen et al. (2013) used a comprehension task where participants listened to several narrations in English with a low proportion of violations. In sentences with no attractor NPs, singular subject agreement violations elicited a bilateral frontal negativity between 150 and 300 ms (interpreted as a LAN) followed by a P600 between 700 and 950 ms. In sentences with complex NPs (e.g., *A catalog with color picture/s sit/\*sits. . .*), those containing number matching attractors elicited an atypical early posterior negativity between 150 and 300 ms, and no P600, while those containing number mismatching attractors elicited neither early posterior negativity nor P600 effects. Although the authors suggest that the posterior negativity resembles the timing and distribution of the N400, the distribution of the negativity related to morphosyntactic violations is rather frontal (and lateralized: LAN) and starts later on (300 ms after the stimulus onset) (see Molinaro et al., 2011, 2015; Tanner, 2015 for an extensive review and discussion). These different ERP patterns might be due to the naturalistic procedure used in this study (i.e., sentences were auditorily presented and embedded in discourse), as compared to the procedure used in most other studies. Regardless of the origin of the atypical early components in this study, the relevant fact for our discussion is that agreement attraction effects reached significance only in ungrammatical sentences (differences between sentences with number matching vs. mismatching attractor NPs), replicating the asymmetric pattern of attraction effects. Shen et al. (2013) interpreted these results as evidence that subject agreement is affected by the presence of number-bearing elements other than the subject itself, with number mismatching elements completely “masking” subject agreement violations.

There are two more studies that explored number attraction effects in subject-verb agreement, but they did not analyze whether these were asymmetric. In an acceptability rating task, Severens et al. (2008) explored in Dutch whether the ambiguity of the determiner of the controller NP affects number attraction. In number match conditions, an atypical ERP pattern related to morphological agreement violations was found, as subject agreement violations only elicited an N400, not followed by a P600. This was interpreted to reflect a blatant violation of the expected verb form during a first, syntactically shallow process that cannot be repaired by further analysis, resulting

in the absence of a P600. In violations involving number mismatching conditions, only a P600 was elicited, which was interpreted as reflecting a deeper syntactic processing triggered by the strong conflict between a shallow syntactic analysis that suggests the first noun (singular) to be the controller and a combinatorial analysis that suggest the noun (plural) agreeing in number with the verb (i.e., the attractor) to be the controller. In other words, the agreement attraction effects were argued to prevent the generation of a N400 component correlated to the ungrammatical verb. Similar findings were reported by Chen et al. (2007) for English singular subject agreement, although this study reported a LAN instead of a N400. Here, a biphasic LAN-P600 pattern was observed in matching conditions, while only a P600 (but no LAN) was reported in mismatching conditions (i.e., *The price of the cars \*were. . .*”).

In summary, the electrophysiological indexes of attraction are mainly reflected by a reduction of main ERP components related to agreement violation detection. The most consistent finding is the reduction of the later P600 component, found in three out of six studies (Kaan, 2002; Tanner et al., 2014, 2016). The other studies showed a reduction of diverse early components: a posterior early negativity (Shen et al., 2013), an N400 (Severens et al., 2008), or a LAN (Chen et al., 2007), but two showed atypical ERP components in the baseline number matching conditions, which might pose problems to the generalizability of attraction effect to other types of dependencies. Further research needs to bring some light on the origin of such heterogeneous patterns of attraction effects in subject-verb agreement. However, it is worth noting that all the studies that explored it found an asymmetrical pattern of attraction effects (Kaan, 2002; Shen et al., 2013; Tanner et al., 2014, 2016), which supports the similarity-based interference account of attraction (Lewis and Vasishth, 2005; Wagers et al., 2009).

## THE PRESENT STUDY

In the present study, we explore for the first time the behavioral and neurophysiological processes of number attraction when processing antecedent-object clitic dependencies in Spanish. We investigate whether antecedent-clitic dependencies are resilient to attraction effects with the aim to provide some experimental evidence on whether clitic dependencies are processed like an agreement dependency or a pronominal dependency.

We carried out two acceptability judgment experiments in Spanish. In each experiment, Spanish native speakers were presented with sentences that had an inanimate object NP containing a PP ([<sub>NP</sub> Det N [<sub>PP</sub> P [<sub>NP</sub> Det N]]]). The Noun inside the PP either matched or mismatched in number with the Noun of the main NP. This complex NP was followed by a left-dislocated object clitic that either matched (grammatical) or mismatched (ungrammatical) in number with the antecedent NP. Clitic left-dislocated structures were investigated for the reason that in peninsular Spanish this is the only way to have the antecedent of the clitic in the same main sentence as the clitic. In this case, all our sentences contained an omitted subject that in its overt form would be placed between the object NP and the

**TABLE 1** | Sample set of experimental items for Experiments 1 and 2.

Sentences	Conditions	
	Attractor number	Grammaticality
<b>Singular objects (Experiments 1 and 2)</b>		
(1) El cartero afirmó que el paquete para el vecino lo entregó a tiempo.	Singular (match)	Grammatical
(2) El cartero afirmó que el paquete para el vecino *los entregó a tiempo. "The postman stated that the package for the neighbor (he) delivered it/*them on time"	Singular (match)	Ungrammatical
(3) El cartero afirmó que el paquete para los vecinos lo entregó a tiempo.	Plural (mismatch)	Grammatical
(4) El cartero afirmó que el paquete para los vecinos *lo entregó a tiempo. "The postman stated that the package for the neighbors (he) delivered it/*them on time"	Plural (mismatch)	Ungrammatical
<b>Plural objects (Experiment 1)</b>		
(5) El cartero afirmó que los paquetes para los vecinos los entregó a tiempo.	Plural (match)	Grammatical
(6) El cartero afirmó que los paquetes para los vecinos *lo entregó a tiempo. "The postman stated that the packages for the neighbors (he) delivered *it/them on time"	Plural (match)	Ungrammatical
(7) El cartero afirmó que los paquetes para el vecino los entregó a tiempo.	Singular (mismatch)	Grammatical
(8) El cartero afirmó que los paquetes para el vecino *lo entregó a tiempo. "The postman stated that the packages for the neighbor (he) delivered *it/them on time"	Singular (mismatch)	Ungrammatical

clitic (see **Table 1**)<sup>1</sup>. In Experiment 1, a self-paced reading task was used and singular and plural antecedent NPs were presented. In Experiment 2, singular antecedent NPs were presented and the acceptability ratings and electrophysiological responses of participants were recorded while reading sentences presented with a RSVP paradigm.

## EXPERIMENT 1

In Experiment 1 we explored whether the number attraction effects previously observed for subject-verb agreement (e.g., Nicol et al., 1997; Pearlmutter et al., 1999; Pearlmutter, 2000; Wagers et al., 2009) obtain during the processing of the dependency between antecedents and clitics. If this dependency is a subtype of agreement as suggested by the *Clitics as Agreement Hypothesis* we expect faster reading times and lower accuracy judging ungrammatical sentences containing antecedent NPs containing a number matching attractor NP (Wagers et al., 2009). Since, for the sake of completeness, we included singular and plural antecedents, we also expect to replicate the number markedness effects of attraction (Nicol et al., 1997; Wagers et al., 2009), so that larger attraction effects (if any) are expected

<sup>1</sup>Note that in these types of structures the object NP (*el paquete para los vecinos*, in the example of **Table 1**) might have been interpreted as a subject which might have predicted the following word to be a verb (*el paquete para los vecinos era...*, "the package for the neighbors was...") instead of a left-dislocated object clitic (*el paquete para los vecinos lo...*, lit. "the package for the neighbors it..."). Due to this ambiguity, one might argue that the NP would only be recognized as an object NP when the verb is reached, and thus participants might be stuck in a garden-path until then. However, since in our materials all the NPs were inanimate, they would have been most likely interpreted as the subjects of an intransitive event, because the subjects of transitive events are more likely to be animate. Thus, reading at the critical region an object clitic instead of a verb would suffice to break the garden-path effect, as it would strongly prioritize interpreting the antecedent NP as an object NP. Importantly, it is worth noting that in the event the antecedent NP was interpreted as a subject NP, the noun inside the PP (*el/los vecino/s*) could not be considered the antecedent of the object clitic (*lo/s*) in Peninsular Spanish. We thank Brian Dillon for pointing out to this possible confound.

for sentences containing singular antecedent NPs, than for sentences containing plural antecedent NPs. However, if clitics establish pronominal dependencies as argued by the *Clitics as Pronouns Hypothesis*, no attraction effects are expected in self-paced readings (Experiment 1), as suggested by previous evidence with other pronominal forms like reflexives (Dillon et al., 2013; Parker and Phillips, 2017). In sum, the presence of attraction effects suggests that antecedent-clitic dependencies are processed as a subtype of agreement.

At this point, we would like to add a cautionary note about the time course at which these effects are to be observed. In most self-paced reading studies, attraction effects appear in the region following the critical verb (Wagers et al., 2009; Lago et al., 2015). In our experimental sentences, as in Pearlmutter et al. (1999), the attractor NP immediately precedes the clitic, so that some attractor number carry-over effects are expected (Wagers et al., 2009). Hence, we expect attraction (and acceptability) effects to arise at the position following the critical word (CW), the clitic.

## Method

### Participants

Sixty native speakers of Spanish (42 females, mean age years 22.7;  $SD = 5.8$ ), undergraduates at the University of the Basque Country (UPV/EHU) were paid for their participation in the study. All Participants gave written informed consent under experimental protocols approved by the Ethics Committee of the UPV/EHU (Comité de Ética para las Investigaciones relacionadas con Seres Humanos, CEISH), in accordance with the Declaration of Helsinki.

### Materials

Experimental materials consisted of 48 sentences. Each sentence had the following structure: a subject NP followed by the main verb and a subordinate clause containing an object NP + object-clitic + subordinate verb + PP (see **Table 1** and Appendix). Crucially, object NPs were third person and contained a singular

or plural head noun and a singular or plural NP inside the modifying PP. Eight experimental conditions were created crossing three factors: Object Number (singular vs. plural) vs. Attractor Number (singular vs. plural) and Grammaticality (grammatical vs. ungrammatical sentences). Each sentence was presented once in each of these conditions.

Additionally, we created 96 filler sentences to introduce some variability in the stimuli. 84 of these filler sentences were grammatical and 12 contained subject-verb agreement violations. We created eight lists containing 144 sentences, from which 48 were experimental sentences (6 per condition) and 96 were fillers. Each list contained a total of 36 (25%) ungrammatical sentences. Each participant was presented with only one of these lists. Each item was presented only once in each list. Four additional sentences (2 grammatical and 2 ungrammatical) were used as practice trials.

## Procedure

Linger (Rohde, 2001) software was used to present the stimuli. Before the experiment started, participants received written instructions about the main procedure. They were asked to read and understand sentences word-by-word as fast as they could by pressing the spacebar in a self-paced reading task. The materials were pseudo-randomized in the following way: no sentences of the same condition were displayed one after another and each experimental sentence (see examples 1–8 in **Table 1**) was followed by a filler sentence. A fixation cross (+) indicated the beginning of each trial. After each sentence a question mark was presented and participants were instructed to press one of two buttons (1 and 2 on the keyboard) depending on whether the previously displayed sentence was grammatical or not. Half of the participants pressed 1 for grammatical sentences and 2 for ungrammatical sentences; the other half used the reversed configuration. All 144 sentences were distributed over 4 blocks, and participants were asked to have short breaks between these blocks. Before the experiment began, participants were familiarized with the procedure by means of a short trial session in which 4 sentences were presented (2 grammatical, 2 ungrammatical sentences). The experiment lasted about 25 min.

## Data Analysis

Acceptability judgment accuracy and reading time data were analyzed with mixed logit and linear mixed effects regression models, respectively. Reading times faster than 50 ms or slower than 4000 ms were excluded (0.5% of the data), and reading times that exceeded a threshold of 2.5 standard deviations by region and condition were excluded (2.3% of the analyzed data). Next, raw reading times were log-transformed to normalize the data and spill-over effects of the previous two words were calculated for each word region<sup>2</sup>.

In the analyses of grammaticality judgment and self-paced reading time data, our binomial variable (whether

a grammaticality judgment was performed or not in the grammaticality judgment task), or log-transformed reading time dependent variables were fitted with (generalized) linear mixed regression models including crossed random and fixed effects (Baayen, 2008; Baayen et al., 2008; Jaeger, 2008). The following sum coded fixed factors were included in the models: Object Number (singular vs. plural), Attractor Number (singular vs. plural), Grammaticality (grammatical vs. ungrammatical), and their interactions. In the reading time analyses, the spill-over effects of the two previous words were also included in the model as fixed effects. When the maximal model failed to converge or showed high correlation parameters between random effects ( $>0.8$ ), we used the backward selection based on  $\chi^2$ . Finally, whenever a significant interaction effect revealed different patterns of results for the involved fixed factors, we run simpler models that split without one of the involved fixed factors in order to find the source of the interaction (e.g., when the three-way interaction was significant, we run two separate models including the Attractor Number, Grammaticality, and their interactions as fixed factors; the maximal random effect structures of the main models was kept). All analyses were carried out in R (version 3.4.0; R Development Core Team, 2013) using the lmerTest package.

## Results

### Grammaticality Judgment Errors (See Table 2)

The maximal random effect structure justified by model comparison included a by-participant Grammaticality random slope. The results showed significant Attractor Number ( $\beta = 0.370$ ,  $SE = 0.070$ ,  $z = 5.224$ ,  $p < 0.001$ ), and Grammaticality ( $\beta = -0.281$ ,  $SE = 0.126$ ,  $z = -2.215$ ,  $p = 0.026$ ) effects. These effects showed that more errors were produced in grammatical than ungrammatical sentences and in sentences where the number of the antecedent NP and the attractor mismatched than matched.

There was also a significant Attractor Number by Object Number interaction ( $\beta = -0.390$ ,  $SE = 0.070$ ,  $z = -5.506$ ,  $p < 0.001$ ), revealing larger attraction effects (more errors judging sentences with number mismatching than matching attractors) in sentences containing singular than plural objects. The simpler models revealed that the attraction effects were only significant in sentences containing a singular object ( $\beta = 0.766$ ,  $SE = 0.106$ ,  $z = 7.168$ ,  $p < 0.001$ ) but not in sentences containing a plural object ( $\beta = -0.022$ ,  $SE = 0.096$ ,  $z = -0.235$ ,  $p < 0.814$ ). Finally, the three-way interaction was marginally significant ( $\beta = -0.130$ ,  $SE = 0.070$ ,  $z = -1.846$ ,  $p < 0.064$ ), revealing different grammaticality by attraction patterns in sentences containing singular and plural objects. The simpler models revealed a non-significant Attractor Number by Grammaticality interaction in sentences containing singular objects ( $\beta = 0.080$ ,  $SE = 0.106$ ,  $z = 0.758$ ,  $p < 0.448$ ), but a significant interaction for sentences with plural objects ( $\beta = -0.218$ ,  $SE = 0.096$ ,  $z = -2.260$ ,  $p < 0.023$ ). The later interaction revealed different direction of attraction effects in grammatical and ungrammatical conditions, but these effects were not significant in either condition (all  $p > 0.10$ ). No further effects were found (all  $z < 2$ ).

<sup>2</sup>The spill-over effects were computed following the steps described in Florian Jaeger's blog: <http://www.hplab.wordpress.com/2008/01/23/modeling-self-paced-reading-data-effects-of-word-length-word-position-spill-over-etc/> and <https://hplab.wordpress.com/2007/11/23/spill-over-effects-in-self-paced-reading/>; accessed June 7, 2017).

**TABLE 2 |** Raw count of errors (from a total of 360 responses per condition; percentages in brackets) and reaction time (ms) values of participants' performance in the grammaticality judgment task in each experimental condition of Experiment 1.

	Grammaticality judgment errors		Response latencies	
	Grammatical	Ungrammatical	Grammatical	Ungrammatical
<b>Singular object</b>				
Singular attractor (match)	22 (6.1%)	15 (4.2%)	591	544
Plural attractor (mismatch)	67 (18.6%)	57 (15.8%)	603	570
Attraction effect	-45 (-12.5%)	-42 (-11.6%)	-11	-26
<b>Plural object</b>				
Plural attractor (match)	38 (10.6%)	34 (9.4%)	583	556
Singular attractor (mismatch)	51 (14.2%)	23 (6.4%)	586	589
Attraction effect	-13 (-3.6%)	11 (3.0%)	-3	-33

**TABLE 3 |** Self-paced reading results (in ms) in each experimental condition of Experiment 1.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	CW	CW+1	CW+2	CW+3
<b>Singular object</b>													
Match-grammatical	391	459	467	435	418	461	438	401	490	450	496	441	703
Match-ungrammatical	385	445	457	428	412	446	426	402	487	480	577	418	555
Grammaticality effect	7	14	10	7	7	16	12	-1	4	-30	-80	23	148
Mismatch-grammatical	382	443	444	416	398	442	433	422	495	485	517	472	715
Mismatch-ungrammatical	401	456	464	410	401	459	428	416	538	516	600	461	608
Grammaticality effect	-19	-12	-20	6	-4	-17	4	6	-42	-31	-84	11	107
Attraction effect (Gramm)	9	15	23	19	21	20	5	-21	-5	-35	-20	-31	-12
Attraction effect (Ungramm)	-16	-11	-6	18	11	-13	-2	-14	-51	-36	-23	-43	-53
<b>Plural object</b>													
Match-grammatical	379	448	454	426	405	467	442	417	524	507	521	472	680
Match-ungrammatical	396	449	482	424	429	480	447	423	511	494	620	433	557
Grammaticality effect	-17	-1	-28	1	-24	-14	-6	-6	13	13	-99	39	123
Mismatch-grammatical	387	447	457	421	422	456	446	411	509	478	512	472	712
Mismatch-ungrammatical	383	438	465	412	410	470	438	412	490	487	612	458	581
Grammaticality effect	4	9	-8	9	12	-14	8	-1	19	-9	-99	14	132
Attraction effect (Gramm)	-8	0	-3	4	-17	11	-5	6	15	29	9	0	-32
Attraction effect (Ungramm)	14	10	17	12	19	11	9	12	21	7	9	-25	-24

R, region; CW, critical word (object-clitic). Region by region means segregated by object number and grammaticality and main grammaticality effects (grammatical minus ungrammatical conditions) and attraction effects (match minus mismatch conditions). Sample sentence (singular object conditions):  $El_{(R1)}$   $cartero_{(R2)}$   $afirmó_{(R3)}$   $que_{(R4)}$   $el_{(R5)}$   $paquete_{(R6)}$   $para_{(R7)}$   $el/los_{(R8)}$   $vecino(s)_{(R9)}$   $lo/los_{(CW)}$   $entregó_{(CW+1)}$   $a_{(CW+2)}$   $tiempo_{(CW+3)}$ .

## Grammaticality Judgment Response Latencies (See Table 2)

None of the effects were significant (all  $t_s < 2$ ).

## Self-paced Reading Response Latencies

The maximal random effect structures justified by model comparison that did not have convergence or high correlation parameter problems did not include any random slopes for regions R5, R6, R7, CW+2, and CW+3 and contained a by-item Attractor Number random slope for regions R8, R9, CW, and CW+1<sup>3</sup> (see Tables 3, 4, for the self-paced reading data, reported in milliseconds, and the mixed-effect model based on

<sup>3</sup>Note that the models with or without the by-item Attractor Number random slope revealed the same patterns of results in all regions (even if the models showed overparameterization in regions R5, R6, R7, CW+2, and CW+3). Analysis with log-transformed residual reading times as a dependent variable also showed similar results as the ones reported with log-transformed reading times.

log-transformed reading times, respectively). The main effect of Object Number was significant at the object region and marginally significant at the following region (R6 and R7) as well as at the two regions after the clitic (CW+1 and CW+2), with slower reading times in sentences containing plural than singular objects. The main effect of Grammaticality was marginally significant at the region after the object noun, which must be random, but most importantly it was fully significant at the region after the clitic (CW+1), revealing that participants were slower reading ungrammatical than grammatical sentences. This Grammaticality effect reversed in the last two regions of the sentence (CW+2 and CW+3). In this regard, the significant Grammaticality by Object Number interaction found at the clitic region revealed that this grammaticality effect was already present at the clitic region in sentences containing singular objects ( $\beta = 0.024$ ,  $SE = 0.008$ ,  $t = 2.773$ ,  $p = 0.005$ ), while it was not significant in sentences containing plural objects ( $p > 0.3$ )



**TABLE 4 |** Linear mixed models for the analysis of the self-paced log-transformed reading times per region in Experiment 1.

Predictor	$\beta$	SE	t-value	p
<b>R5</b>				
(Intercept)	2.897	0.115	25.132	<0.001
Object number	0.007	0.005	1.454	0.146
Attractor number	-0.003	0.005	-0.593	0.553
Grammaticality	0.005	0.005	1.118	0.263
Attractor number × Object number	0.003	0.005	0.536	0.591
Grammaticality × Object number	0.002	0.005	0.323	0.746
Attractor number × Grammaticality	-0.006	0.005	-1.243	0.213
3-way interaction	-0.012	0.005	-2.439	0.014
Spillover_1	0.292	0.017	16.867	<0.001
Spillover_2	0.219	0.015	14.910	<0.001
<b>R6</b>				
(Intercept)	2.094	0.144	14.556	<0.001
Object number	0.013	0.006	2.347	0.019
Attractor number	-0.002	0.006	-0.394	0.694
Grammaticality	0.002	0.006	0.331	0.741
Attractor number × Object number	-0.007	0.006	-1.284	0.199
Grammaticality × Object Number	0.003	0.006	0.471	0.637
Attractor number × Grammaticality	0.008	0.006	1.442	0.149
3-way interaction	-0.002	0.006	-0.271	0.786
Spillover_1	0.387	0.021	18.541	<0.001
Spillover_2	0.276	0.021	13.300	<0.001
<b>R7</b>				
(Intercept)	3.094	0.112	27.639	<0.001
Object number	0.008	0.005	1.726	0.084
Attractor number	0.005	0.005	1.110	0.267
Grammaticality	-0.009	0.005	-1.896	0.058
Attractor number × Object Number	-0.005	0.005	-1.018	0.308
Grammaticality × Object Number	0.003	0.005	0.632	0.527
Attractor number × Grammaticality	-0.001	0.005	-0.274	0.784
3-way interaction	-0.003	0.005	-0.568	0.569
Spillover_1	0.213	0.016	13.747	<0.001
Spillover_2	0.275	0.019	14.755	<0.001
<b>R8</b>				
(Intercept)	3.123	0.106	29.527	<0.001
Object number	0.003	0.004	0.732	0.464
Attractor Number	0.004	0.005	0.898	0.373
Grammaticality	0.002	0.004	0.649	0.516
Attractor number × Object number	-0.010	0.004	-2.286	0.022
Grammaticality × Object number	0.002	0.004	0.556	0.578
Attractor number × Grammaticality	-<0.001	0.004	-0.149	0.881
3-way interaction	0.001	0.004	0.285	0.775
Spillover_1	0.254	0.016	15.323	<0.001
Spillover_2	0.217	0.014	15.502	<0.001
<b>R9</b>				
(Intercept)	1.913	0.165	11.538	<0.001
Object number	-0.002	0.006	-0.320	0.749
Attractor number	0.003	0.006	0.486	0.629
Grammaticality	-<0.001	0.006	-0.047	0.962
Attractor number × Object number	-0.019	0.006	-3.052	0.002
Grammaticality × Object number	-0.013	0.006	-2.149	0.031
Attractor number × Grammaticality	0.004	0.006	0.638	0.523

(Continued)

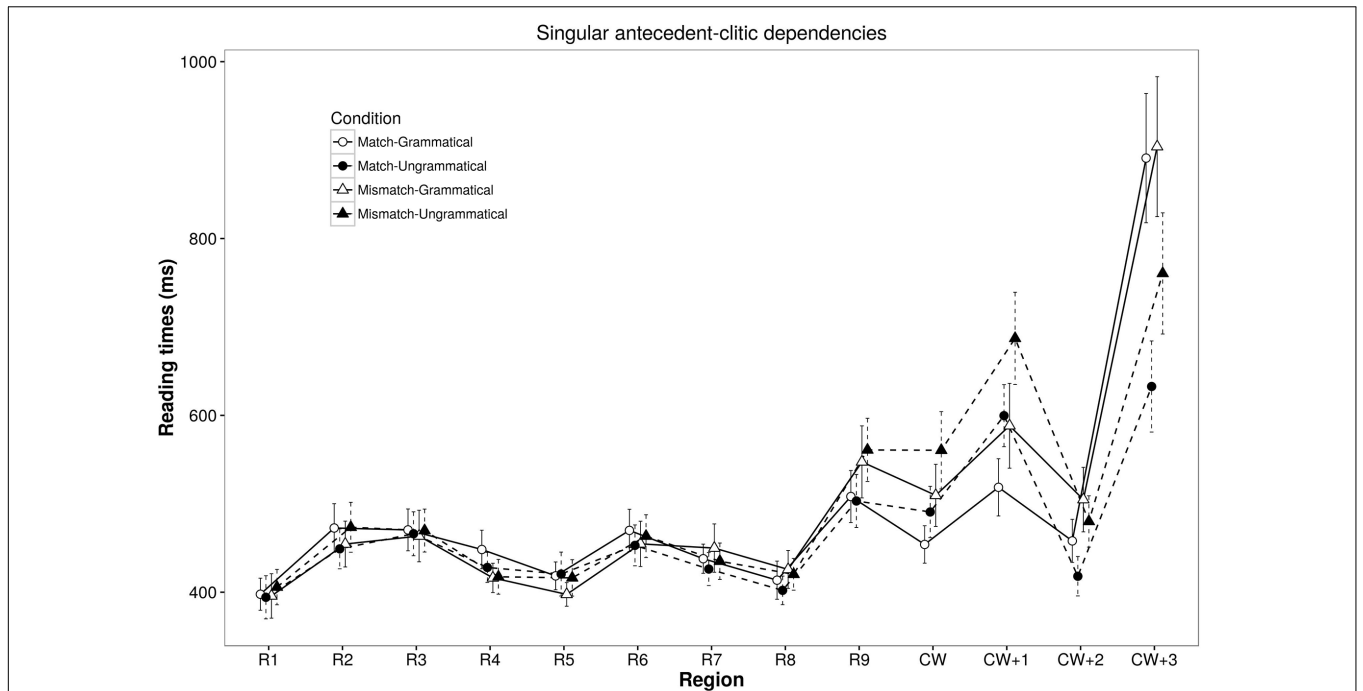
TABLE 4 | Continued

Predictor	$\beta$	SE	t-value	p
3-way interaction	-0.003	0.006	-0.599	0.549
Spillover_1	0.356	0.024	14.636	<0.001
Spillover_2	0.350	0.023	15.086	<0.001
<b>CW</b>				
(Intercept)	3.469	0.147	23.547	<0.001
Object number	0.008	0.006	1.318	0.187
Attractor number	0.008	0.006	1.214	0.230
Grammaticality	0.008	0.006	1.315	0.188
Attractor number $\times$ Object number	-0.023	0.006	-3.645	<0.001
Grammaticality $\times$ Object number	-0.016	0.006	-2.629	0.008
Attractor number $\times$ Grammaticality	0.001	0.006	0.260	0.794
3-way interaction	-<0.001	0.006	-0.039	0.968
Spillover_1	0.189	0.017	10.884	<0.001
Spillover_2	0.249	0.024	10.253	<0.001
<b>CW+1</b>				
(Intercept)	3.456	0.159	21.618	<0.001
Object number	0.014	0.007	1.895	0.058
Attractor number	0.020	0.008	2.563	0.013
Grammaticality	0.070	0.007	9.256	<0.001
Attractor number $\times$ Object number	-0.001	0.007	-0.236	0.813
Grammaticality $\times$ Object number	0.004	0.007	0.571	0.568
Attractor number $\times$ Grammaticality	-0.003	0.007	-0.450	0.652
3-way interaction	-0.006	0.007	-0.854	0.393
Spillover_1	0.188	0.021	8.562	<0.001
Spillover_2	0.269	0.020	13.159	<0.001
<b>CW+2</b>				
(Intercept)	5.283	0.142	37.285	<0.001
Object number	0.013	0.007	1.810	0.070
Attractor number	0.026	0.007	3.631	<0.001
Grammaticality	-0.035	0.007	-4.828	<0.001
Attractor number $\times$ Object number	-0.014	0.007	-1.930	0.053
Grammaticality $\times$ Object number	0.004	0.007	0.594	0.552
Attractor number $\times$ Grammaticality	0.011	0.007	1.506	0.132
3-way interaction	-0.006	0.007	-0.834	0.404
Spillover_1	0.015	0.016	0.926	0.354
Spillover_2	0.111	0.020	5.573	<0.001
<b>CW+3</b>				
(Intercept)	2.620	0.217	12.060	<0.001
Object number	-0.008	0.010	-0.820	0.412
Attractor number	0.011	0.010	1.103	0.270
Grammaticality	-0.136	0.010	-13.249	<0.001
Attractor number $\times$ Object number	-0.002	0.010	-0.178	0.858
Grammaticality $\times$ Object number	-0.005	0.010	-0.542	0.588
Attractor number $\times$ Grammaticality	0.009	0.010	0.876	0.380
3-way interaction	-0.017	0.010	-1.719	0.085
Spillover_1	0.375	0.027	13.918	<0.001
Spillover_2	0.250	0.023	10.720	<0.001

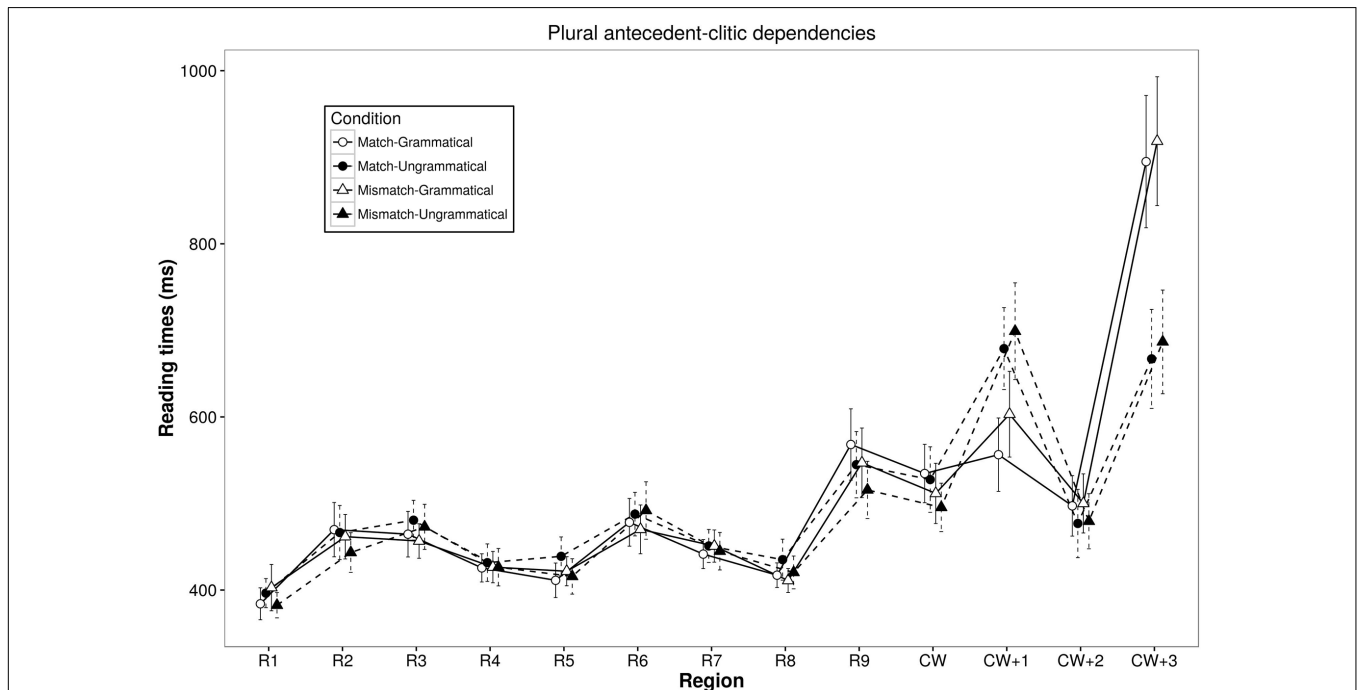
R, region; CW, critical word (clitic). Sample sentence for the analyzed regions (singular object conditions): (. . .) el<sub>(R5)</sub> paquete<sub>(R6)</sub> para<sub>(R7)</sub> el/los<sub>(R8)</sub> vecino(s)<sub>(R9)</sub> lo/\*los<sub>(CW)</sub> entregó<sub>(CW+1)</sub> a<sub>(CW+2)</sub> tiempo<sub>(CW+3)</sub>.

(see **Figures 1, 2**). This interaction was also found in the region preceding the clitic region where violations might occur (R9), and it only signals the presence of non-significant random trends of

opposite effects of grammaticality in sentences with singular vs. plural objects (both  $ps > 0.1$ ). This effect must be random and is not further discussed.



**FIGURE 1 |** Self-paced reading results of sentences with *singular object nouns* (Experiment 1). Region by region means segregated by object noun number and grammaticality. The bars associated with each mean represent standard errors. Sample sentence: *El*<sub>(R1)</sub> *cartero*<sub>(R2)</sub> *afirmó*<sub>(R3)</sub> *que*<sub>(R4)</sub> *el*<sub>(R5)</sub> *paquete*<sub>(R6)</sub> *para*<sub>(R7)</sub> *el/los*<sub>(R8)</sub> *vecino(s)*<sub>(R9)</sub> *lo/\*los*<sub>(CW)</sub> *entregó*<sub>(CW+1)</sub> *a*<sub>(CW+2)</sub> *tiempo*<sub>(CW+3)</sub>.



**FIGURE 2 |** Self-paced reading results of sentences with *plural object nouns* (Experiment 1). Region by region means segregated by object noun number and grammaticality. The bars associated with each mean represent standard errors. Sample sentence: *El*<sub>(R1)</sub> *cartero*<sub>(R2)</sub> *afirmó*<sub>(R3)</sub> *que*<sub>(R4)</sub> *los*<sub>(R5)</sub> *paquetes*<sub>(R6)</sub> *para*<sub>(R7)</sub> *el/los*<sub>(R8)</sub> *vecino(s)*<sub>(R9)</sub> */\*lo/los*<sub>(CW)</sub> *entregó*<sub>(CW+1)</sub> *a*<sub>(CW+2)</sub> *tiempo*<sub>(CW+3)</sub>.

The main effect of Attractor Number was significant at the post-clitic region and the following region (CW+1 and CW+2), showing that participants were slower reading sentences containing attractors that mismatched rather than matched in number with the object.

The Attractor Number by Object Number interaction was significant at the regions of the determiner of the attractor, the attractor and the clitic (R8, R9, CW), as well as marginally significant two regions after the clitic (CW+2). This interaction seems to reveal plural marking slow down effects (and carry-over of such effects) rather than number attraction effects. This is because, in the case of sentences containing singular objects, participants showed slower reading times in sentences containing number mismatching plural rather than number matching singular attractors (R8:  $\beta = 0.014$ ,  $SE = 0.007$ ,  $t = 2.096$ ,  $p = 0.041$ ; R9:  $\beta = 0.022$ ,  $SE = 0.010$ ,  $t = 2.196$ ,  $p = 0.033$ ; CW:  $\beta = 0.024$ ,  $SE = 0.008$ ,  $t = 2.773$ ,  $p < 0.001$ ; CW+2:  $\beta = 0.037$ ,  $SE = 0.009$ ,  $t = 3.941$ ,  $p < 0.001$ ). However, in sentences with plural objects, no attractor number effects were found in regions R8, CW and CW+2 (all  $ps > 0.10$ ), and a marginally significant reversed attraction effect was only found at the region preceding the clitic (R9:  $\beta = -0.015$ ,  $SE = 0.009$ ,  $t = -1.725$ ,  $p = 0.085$ ), with faster reading times in sentences containing number mismatching singular rather than number matching plural attractors.

The Grammaticality by Attractor Number two-way interaction was not significant at any region. The three-way interaction was significant at R5, which must have been random, and was marginally significant at the last region at which wrap-up effects occur.

## Discussion

The grammaticality judgment accuracy data replicated two of the most common findings in agreement: (a) *An attraction effect*: participants produced more grammaticality judgment errors when sentences contained an attractor that mismatched the number of the antecedent NP as compared to sentences containing a number matching attractor (Nicol et al., 1997; Franck et al., 2015); (b) *A markedness effect*: attraction effects obtained with singular but not with plural antecedent NPs (Bock and Miller, 1991; Eberhard, 1997; Pearlmutter et al., 1999; Pearlmutter, 2000; Wagers et al., 2009). Plural attractors disrupted participants' grammaticality judgment accuracy both when accepting grammatical sentences and when rejecting ungrammatical ones. This replicates the finding of attraction effects in the judgment of grammatical sentences by Nicol et al. (1997: Experiment 2), in contrasts with the results of Franck et al. (2015: Experiment 3).

Reading time results are less conclusive. This is because, despite the inclusion of spill-over effects in the model, the main number attraction effects seem to reflect carry-over effects of the larger difficulty of processing the number of the plural attractor presented just before the clitic (Pearlmutter et al., 1999; Wagers et al., 2009): in sentences containing singular antecedent NPs, the presence of plural attractors, as compared to singular ones, slowed down participants' reading times. This effect persisted at the clitic region and the following ones, both in grammatical and

ungrammatical sentences. In contrast, in sentences containing plural antecedent NPs, reading times at the attractor noun region were slower for matching plural than mismatching singular attractors. These slow down effects occurred only in grammatical sentences, and persisted only until the following clitic region. The fact that these effects appear in sentences with singular and plural antecedent NPs at the regions where the attractor is presented suggests that part, if not all, of the attractor number effects are due to the greater reading and processing cost of morphologically marked plural attractors. Consequently, we argue that these effects are not *bona fide* agreement attraction effects.

Importantly, similar grammaticality effects obtained while reading sentences containing singular or plural antecedent object NPs. In both cases, the presence of clitics mismatching in number with their antecedent NP (ungrammatical) led to slower reading times than those obtained for matching antecedent-clitic pairs (grammatical). These differences arose at the clitic region in sentences with singular dependencies, and at the following region in sentences with plural dependencies.<sup>4</sup> But in both cases, at the two-final regions ungrammatical sentences were read faster than grammatical ones. This is probably because, once participants detected the ungrammaticality of the sentence at previous regions (CW or CW+1), they simply speeded up reading the sentence to complete the grammaticality judgment task.

Grammaticality judgment accuracy data indicate that attraction effects occur both in grammatical and ungrammatical sentences in antecedent-clitic dependencies, but no significant effects were found in reading times. Reading time measures revealed no attraction effects, and we argue this is because they were obscured by the carry-over effects of the processing of the preceding plural attractor NPs. Hence, accuracy data suggests that clitic dependencies are affected by the same factors as subject-verb agreement, which would favor the *Clitics as Agreement Hypothesis* (e.g., Suñer, 1988; Franco, 2000) according to which clitics are agreement morphemes. In contrast, reading time data suggest that the processing of clitic dependencies is not affected by the same factors as subject-verb agreement and that might be processed differently (as suggested by Phillips et al., 2011; Dillon et al., 2013), which could be interpreted as evidence favoring the *Clitics as Pronouns Hypothesis*, according to which clitics are pronouns generated at the argument position that moved to the verb (e.g., Kayne, 1975; Torrego, 1988; Uriagereka, 1995; Sportiche, 1998; Anagnostopoulou, 2003; Marchis and Alexiadou, 2013, among others). In order to shed more light on this issue, in Experiment 2 we use ERP methods with which, due to their finer temporal resolution than self-paced reading methods, we might be able to detect the presence of (any) attraction effects that overcome the carry-over effects of the processing of plural attractor nouns.

<sup>4</sup>The fact that grammaticality effects were already detected at the critical region in singular object antecedent conditions suggests that the possible garden path effects elicited while reading the ambiguous antecedent NP were solved at the critical object clitic region, without the need to wait until the following verb region (see Footnote 1).

## EXPERIMENT 2

Self-paced reading measures in Experiment 1 did not reveal attraction effects. Due to the finer temporal resolution of electrophysiological measures, in Experiment 2 we sought to detect attraction effects, if there are any, at clitic position. In this case, and following previous ERP evidence (Rossi et al., 2014), we expect clitic number violations to elicit a P600 component, which might also be preceded by a negative (N400 or LAN) component similar to the one reported for gender violations (Silva-Pereyra et al., 2012). Importantly, if clitics are agreement morphemes, we should be able to detect similar number attraction effects as those reported for subject-verb agreement (Kaan, 2002; Shen et al., 2013; Tanner et al., 2014, 2016), and we expect attraction effects to reduce the magnitude of the ERP components, particularly the P600.

### Method

#### Participants

Forty-six native speakers of Spanish (mean age 21.96 years;  $SD = 5.29$ ), undergraduates at the University of the Basque Country (UPV/EHU), were paid for their participation. All participants were right-handed (Edinburgh Handedness Inventory, Oldfield, 1971) and they had normal or corrected to normal vision. All participants gave written informed consent under an experimental protocol approved by the Ethics Committee of the UPV/EHU (CEISH), in accordance with the Declaration of Helsinki.

#### Materials and Procedure

We used the same materials as in Experiment 1. However, in order to simplify the experimental design, only singular clitic dependencies were tested, and only four experimental conditions created, crossing two factors: Attractor Number (singular vs. plural) and Grammaticality (grammatical vs. ungrammatical object-clitics; see **Table 1**, examples 1–4). We created four lists containing 168 sentences, from which 48 were experimental sentences (12 per condition) and 120 were fillers (24 contained singular subject-verb agreement violations and 96 were grammatical sentences). Thus, only 28.6% of the sentences were ungrammatical (48 out of 168). All further details were the same as in Experiment 1.

The experiment was performed using Presentation® software (Version 16.0<sup>5</sup>). Before the experiment started, participants were instructed about the EEG procedure and seated comfortably in a quiet room in front of a 17 inch monitor. All sentences were displayed in the middle of the screen word-by-word for 350 ms ( $ISI = 200$  ms) in a rapid serial visual presentation paradigm. Materials were pseudo-randomized in the following way: no sentences of the same condition were displayed one after another and each experimental sentence was followed by a filler sentence. A fixation cross (+) indicated the beginning of each trial. After each sentence the words *CORRECTO* ('correct') and *INCORRECTO* ('incorrect') appeared on screen for 3000 ms, asking subjects to press one of two buttons (left or right, with

response hand counterbalanced across participants) depending on whether the previously displayed sentence was grammatical or not. All 168 sentences were distributed over four blocks. Participants could take short breaks between blocks. Before the experiment began, participants ran a four trial procedure familiarization session. They were instructed not to blink or move when sentences were displayed and to make the grammaticality judgment as fast as possible. The whole session lasted no longer than 1 h.

#### EEG Recording

The ERPs were recorded from 32 scalp electrodes mounted in an elastic cap (Electro-Cap International, Inc.; 10–20 system). The electrodes were placed as follows: Fp1, Fpz, Fp2, F7, F3, Ground electrode, FZ, F4, F8, C5A, C1A, C2A, C6A, T3, C3, CZ, C4, T4, TCP1, C1P, C2P, TCP2, T5, P3, PZ, P4, T6, P1P, P2P, O1, Oz, and O2. All electrodes were referenced to left and right mastoids and rereferenced off-line to the nasal-bone electrode. The vertical and horizontal electro-oculograms (VEOG and HEOG) were recorded from electrodes located below (VEOG) and at the outer canthus (HEOG) of the right eye. Electrode impedance was kept below 10 k $\Omega$ . The electrical signals were digitalized on-line at a rate of 250 Hz and filtered off-line with a bandpass of 0.1–35 Hz (half-amplitude cut-offs). After the stimuli were recorded, the artifact rejection procedure was applied (off-line) in order to exclude periods containing eye blinks, head movements or technical artifacts from the data analysis.

#### Data Analysis

The same type of analysis as in Experiment 1 was performed for the behavioral data analysis, with the difference that models only included participants as a unique random effect (item random effect could not be added due to coding limitations in the ERP experimental design). The maximal random effect structure justified by model comparison included a by-participant Grammaticality random slope in the error and response latency analyses. For the electrophysiological data, ANOVA analyses were performed. Average ERPs were computed for each word and each electrode and the 200 ms pre-stimulus baseline was used. Trials with artifacts were excluded from averages. For statistical analyses 9 regions of interest (ROI) were generated, 6 for lateral and 3 for midline electrodes: left frontal (F7, F3, C5A), left central (T3, C3, TCP1), left parietal (T5, P3, O1), right frontal (F4, F8, C6A), right central (C4, T4, TCP2) and right parietal (P4, T6, O2). Midline electrodes were analyzed separately and three ROIs were created for them: frontal (C1A, FZ, C2A), central (C1P, Cz, C2P) and parietal (P1P, Pz, P2P).

As for lateral electrodes, an overall ANOVA was performed for the four within-subject variables included in the analyses: Attractor Number (singular vs. plural), Grammaticality (grammatical vs. ungrammatical), Hemisphere (left vs. right) and Region (frontal vs. central vs. posterior). Midline electrodes analysis included Region (central frontal vs. central vs. central posterior), Attractor Number (singular vs. plural), and Grammaticality (grammatical vs. ungrammatical), and they were analyzed separately from lateral electrodes. Further statistical analyses (MANOVAs) were conducted for each

<sup>5</sup>www.neurobs.com

particular ROI whenever appropriate. Effects for Hemisphere or Region factors were only reported when they interacted with any of the main experimental manipulations: Attractor Number and Grammaticality.

Since ERPs are very sensitive to differences in the context preceding the critical region, our main analysis focused on the ERP components elicited by grammaticality effects in sentences containing singular and plural attractors, separately (examples 1 vs. 2; and 3 vs. 4 in **Table 1**, respectively). However, in order to explore asymmetric grammatical effects on attraction (Wagers et al., 2009; Shen et al., 2013), we also compared the main number attraction effects in grammatical and ungrammatical sentences separately (examples 1 vs. 3 and 2 vs. 4 in **Table 1**).

## Results

### Grammaticality Judgment Errors

Analyses revealed significant Attractor Number ( $\beta = 0.626$ ,  $SE = 0.074$ ,  $z = 8.403$ ,  $p < 0.001$ ), with more errors produced in sentences containing number mismatching plural attractors than in sentences containing number matching singular attractors. The significant Attractor Number by Grammaticality interaction ( $\beta = 0.188$ ,  $SE = 0.074$ ,  $z = 2.528$ ,  $p = 0.011$ ) revealed that attraction effects were larger in ungrammatical sentences ( $\beta = 0.814$ ,  $SE = 0.104$ ,  $z = 7.803$ ,  $p < 0.001$ ) than in grammatical ones ( $\beta = 0.438$ ,  $SE = 0.106$ ,  $z = 4.111$ ,  $p < 0.001$ ; see **Table 5**).

### Grammaticality Judgment Response Latencies

The main effects of Attractor Number ( $\beta = 0.048$ ,  $SE = 0.012$ ,  $t = 3.975$ ,  $p < 0.001$ ) and Grammaticality were significant ( $\beta = -0.117$ ,  $SE = 0.018$ ,  $t = -6.499$ ,  $p < 0.001$ ), showing that participants were faster judging sentences containing number mismatching than matching attractors and were also faster rejecting ungrammatical sentences than accepting grammatical ones. The significant Attractor Number by Grammaticality interaction ( $\beta = 0.042$ ,  $SE = 0.012$ ,  $t = 3.458$ ,  $p < 0.001$ ) revealed attraction effects only when ungrammatical sentences had to be rejected ( $\beta = 0.095$ ,  $SE = 0.017$ ,  $t = 5.361$ ,  $p < 0.001$ ; with slower responses for sentences containing number mismatching than matching attractors).

### ERP Results

Based on visual inspection and on previous ERP studies (Kaan, 2002; Chen et al., 2007; Severens et al., 2008; Dillon et al., 2013; Shen et al., 2013; Tanner et al., 2014, 2016), three main time windows were chose for statistical analyses at the *Clitic region*: 300–500 ms; 500–700 ms; and 700–900 ms.

### 300–500 ms time window

At both the lateral and the midline electrodes, the Attractor Number by Grammaticality by Region three-way interactions were significant Lateral:  $F(2,90) = 5.59$ ;  $p = 0.015$ ; Midline:  $F(2,90) = 7.83$ ;  $p = 0.004$ . To better understand this interaction we conducted follow-up analyses examining the mean Grammaticality effects in sentences containing number matching singular and number mismatching plural attractors separately at each ROI (see **Figure 3**). In sentences with number matching singular attractors, a larger negativity was found over frontal (but no central or posterior) sites of the scalp for ungrammatical sentences in both lateral electrodes [Frontal:  $F(1,45) = 10.44$ ;  $p = 0.002$ ; Central and Posterior: both  $ps > 0.1$ ], and midline electrodes [Frontal:  $F(1,45) = 7.39$ ,  $p = 0.009$ ; Central and Posterior: both  $ps > 0.1$ ] as compared to grammatical sentences. No statistically significant Grammaticality effect obtained in sentences with number mismatching plural attractors at any region, neither in lateral nor midline electrodes.

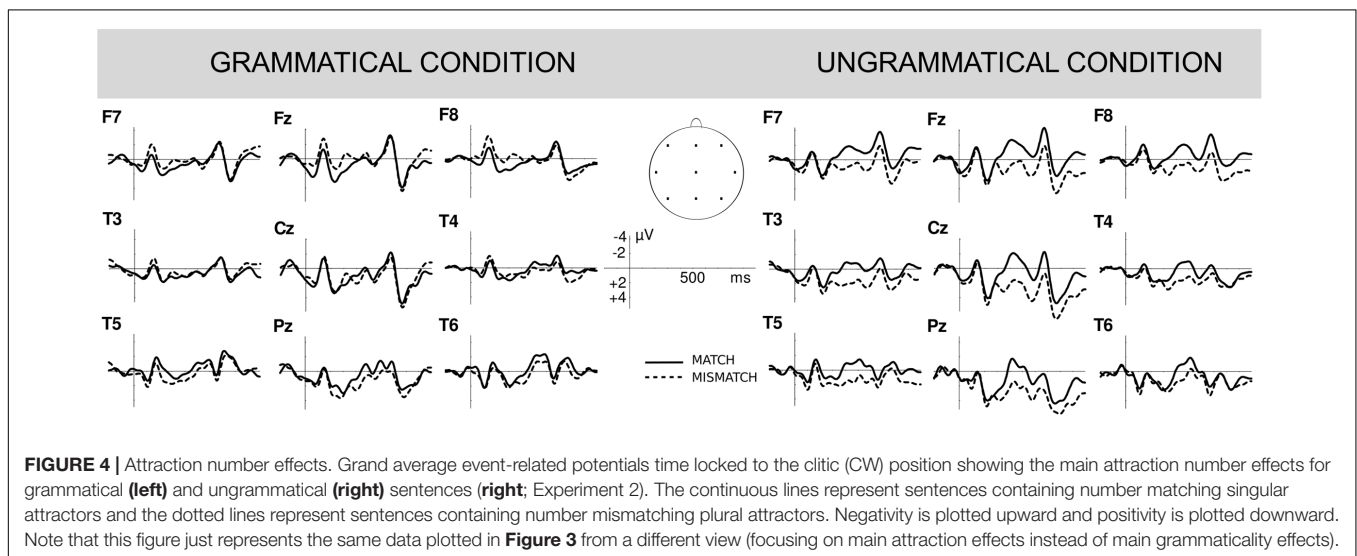
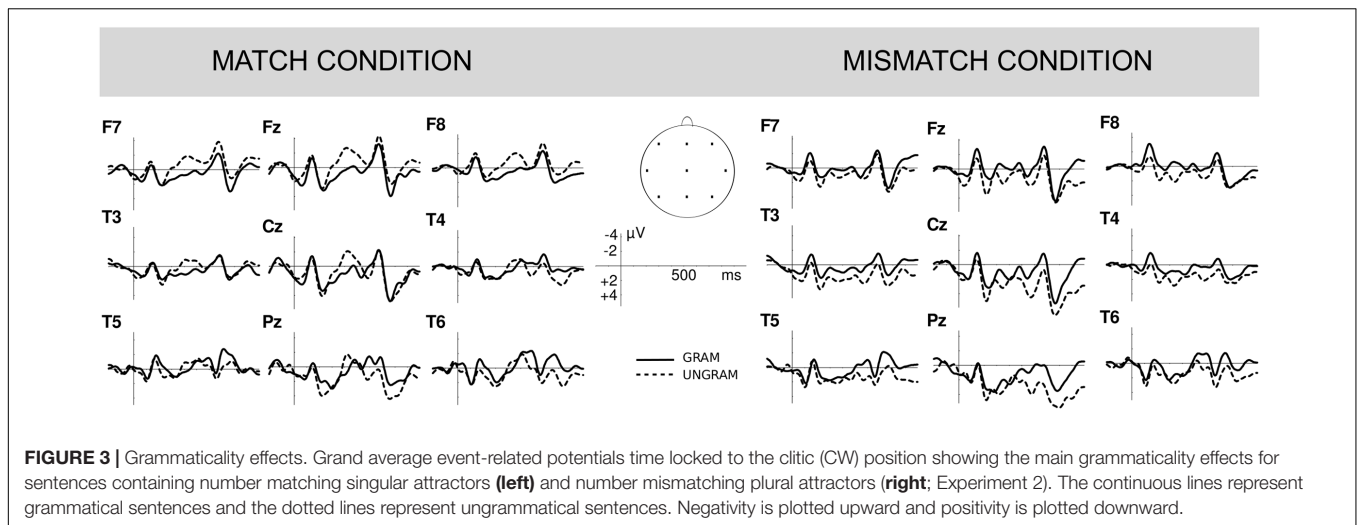
So far we focused on the main grammaticality effects elicited by sentences containing number matching and number mismatching attractors. However, in order to separately assess the presence of an asymmetrical grammaticality of attraction effect, we also conducted *complementary analyses* that focused on Attractor Number effects (see **Figure 4**). No statistically significant attraction effects obtained in grammatical sentences at any region, in either lateral or midline electrodes (all  $p$ -values  $> 0.05$ ). However, significant attraction effects obtained in ungrammatical sentences, with larger negativity in number matching than in number mismatching conditions over all regions, both in lateral electrodes [Frontal:  $F(1,45) = 8.40$ ,  $p = 0.006$ ; Central:  $F(1,45) = 8.42$ ,  $p = 0.006$ ; Posterior:  $F(1,45) = 6.55$ ;  $p = 0.014$ ], and in midline electrodes [Frontal:  $F(1,45) = 9.11$ ,  $p = 0.004$ ; Central:  $F(1,45) = 10.16$ ,  $p = 0.003$ ; Posterior:  $F(1,45) = 6.81$ ,  $p = 0.012$ ].

### 500–700 ms time window

At the lateral and midline electrodes, the two-way interaction of Grammaticality by Region was significant at 500–700 ms [Lateral:  $F(2,90) = 12.46$ ,  $p < 0.001$ ; Midline:  $F(2,90) = 7.26$ ,  $p = 0.006$ ] indicating a different electrophysiological response to grammatical vs. ungrammatical stimuli over frontal, central and posterior sites of the scalp. To better understand this interaction we conducted follow-up analyses examining the mean Grammaticality effects over the different regions of the scalp. This analysis showed a larger positivity for

**TABLE 5 |** Raw count of errors (from a total of 548 responses per condition; percentages in brackets) and reaction time (ms) values of participants' performance in the grammaticality judgment task in each experimental condition of Experiment 2.

	Grammaticality judgment errors		Response latencies	
	Grammatical	Ungrammatical	Grammatical	Ungrammatical
<b>Singular object</b>				
Singular attractor (match)	36 (5.8%)	37 (6.0%)	749	564
Plural attractor (mismatch)	77 (13.9%)	128 (27.4%)	775	699
Attraction effect	-41 (8.1%)	-91 (21.4%)	-26	-134



ungrammatical sentences than grammatical ones over central and posterior, but non-significant effects at frontal sites [Lateral electrodes: Frontal:  $F(1,45) = 0.10$ ,  $p = 0.894$ ; Central:  $F(1,45) = 4.04$ ,  $p = 0.051$ ; Posterior:  $F(1,45) = 6.66$ ,  $p = 0.013$ ; Midline electrodes: Frontal:  $F(1,45) = 1.19$ ,  $p = 0.282$ ; Central  $F(1,45) = 2.52$ ,  $p = 0.120$ , Posterior:  $F(1,45) = 4.18$ ,  $p = 0.047$ ]. In addition, a main effect of Attractor Number was observed, with larger negativity over all electrode sites for sentences containing number matching singular attractors vs. number mismatching plural attractors [Lateral:  $F(1,45) = 5.30$ ,  $p = 0.026$ ; Midline:  $F(1,45) = 5.05$ ,  $p = 0.030$ ]. None of the interactions involving the Attractor Number factor yielded significance, suggesting that the distribution of the significant main grammaticality effects reported above were similar for sentences containing number matching singular attractors [Grammaticality  $\times$  Region in Lateral:  $F(2,90) = 9.47$ ,  $p = 0.002$ ; and Midline electrodes:  $F(2,90) = 6.86$ ,  $p = 0.007$ ] and number mismatching plural attractors [Grammaticality  $\times$  Region in Lateral electrodes:  $F(2,90) = 3.81$ ,  $p = 0.049$ ; and

Grammaticality effect in Midline electrodes:  $F(1,45) = 6.52$ ,  $p = 0.016$ ].

For the sake of completeness, we also performed *complementary analyses* to separately examine the mean Attractor Number effects in grammatical and ungrammatical sentences. In grammatical sentences, none of the effects approached significance at any site (all  $F_s < 1$ ). In contrast, in ungrammatical sentences a main effect of Attractor Number was found over the lateral and midline sites [Lateral  $F(1,45) = 5.43$ ,  $p = 0.024$ ; midline:  $F(1,45) = 5.56$ ,  $p = 0.023$ ], revealing that the larger negativity elicited by number matching singular attractors vs. number mismatching plural ones at the 300–500 ms time window continued to be significant at the 500–700 ms time window. This effect was no longer significant at the 700–900 ms time window (Lateral and Midline: both  $p_s > 0.1$ ).

#### 700–900 ms time window

At lateral and midline electrodes, the same Grammaticality by Region interaction pattern reported for the 500–700 ms time

window was found [Lateral:  $F(2,90) = 16.38, p < 0.001$ ; Midline:  $F(2,90) = 11.74, p = 0.001$ ]. The analyses of Grammaticality effects replicated the pattern reported in the 500–700 ms time window: [Lateral electrodes: Frontal:  $F(1,45) = 0.27, p = 0.609$ ; Central:  $F(1,45) = 3.43, p = 0.07$ ; Posterior:  $F(1,45) = 7.42, p = 0.009$ ; Midline electrodes: Frontal:  $F(1,45) = 0.39, p = 0.536$ ; Central  $F(1,45) = 1.81, p = 0.185$ ; Posterior:  $F(1,45) = 5.01, p = 0.030$ ]. However, in the 700–900 ms time window the main effect of Attractor Number was not significant [Lateral:  $F(1,45) = 1.43, p = 0.237$ ; Midline:  $F(1,45) = 1.79, p = 0.188$ ]. None of the interactions involving Attractor Number factor yielded significance, indicating similar distribution of the grammaticality effects reported above in sentences containing number matching singular [Grammaticality  $\times$  Region in Lateral:  $F(2,90) = 11.31, p = 0.001$ ; and Midline electrodes:  $F(2,90) = 8.20, p = 0.003$ ] and number mismatching plural attractors [Grammaticality  $\times$  Region in Lateral:  $F(2,90) = 5.343, p = 0.020$ ; and Midline electrodes:  $F(2,90) = 4.08, p = 0.039$ ].

Finally, none of the *complementary analyses* focused on examining the mean Attractor Number effects in grammatical and ungrammatical sentences yielded significance at any site (all  $ps > 0.1$ ).

## GENERAL DISCUSSION

In two experiments, we investigated the effects of number attraction on Spanish object clitic dependencies, elicited by number mismatching attractor NPs intervening between the clitic and its antecedent. In Experiment 1, grammaticality judgment accuracy data revealed number attraction effects and number markedness effects, since attraction effects were detected only when the antecedent-clitic dependency was singular, replicating the number markedness effect reported in agreement dependencies (Bock and Miller, 1991; Eberhard, 1997; Pearlmutter et al., 1999; Pearlmutter, 2000; Wagers et al., 2009). However, on-line reading times failed to reveal attraction effects, possibly because of the greater carry-over effect of the slow down originated while reading the plural attractor NPs. In Experiment 2, number attraction effects were detected both by grammaticality judgment data and electrophysiological measures. Grammaticality judgment accuracy and response time data revealed number attraction effects in antecedent-clitic dependency resolution, since there were more errors and slower RTs in sentences containing number mismatching attractors vs. number matching ones. Additionally, asymmetrical attraction was observed, that is, attraction effects were larger for ungrammatical sentences than for grammatical ones (replicating Franck et al., 2015: Experiment 3). As discussed next, these patterns of results were also replicated by the ERP data.

### Electrophysiological Indexes of Antecedent-Clitic Dependencies and Number Attraction

In Experiment 2, violations in sentences with singular attractors (e.g., *... el paquete para el vecino \*los...*) elicited a frontal negativity followed by a P600 component. These components

have been previously reported for antecedent-clitic dependency violations, but not simultaneously: Silva-Pereyra et al. (2012) report an N400 for feminine gender violation and a P600 for masculine gender violation, while Rossi et al. (2014) report a P600 for both gender and number violations. Our biphasic ERP pattern replicates the one usually reported for agreement violations (see Molinaro et al., 2011) and other types of pronominal dependency violations such as reflexives or subject pronouns (Schmitt et al., 2002; Hammer et al., 2005, 2008; Lamers et al., 2006).

Regarding number attraction effects, violations involving singular antecedents and plural clitics with intervening plural attractors elicited a P600 component with no trace of a preceding negativity (e.g., *el paquete para los vecinos \*los...*). This pattern of results, together with those from grammaticality judgment accuracy data in Experiments 1 and 2, reveals greater difficulty detecting clitic number violations when a mismatching plural attractor intervenes. We interpret the absence of a negative component as signaling an attraction effect due to the mismatching attractor (replicating Chen et al., 2007; Severens et al., 2008; Shen et al., 2013). We did not find a reduction of the amplitude of the P600 component that could be interpreted as evidence for attraction effects, as in some studies on agreement (Kaan, 2002; Shen et al., 2013; Tanner et al., 2014).

Importantly, our results also revealed electrophysiological indexes of *asymmetrical* attraction effects: attraction effects only occurred in ungrammatical sentences, not in grammatical ones. In ungrammatical sentences, plural clitics with singular antecedents elicited a large and broadly distributed negativity when preceded by plural attractors, as compared to those preceded by number matching singular attractors. No equivalent differences were found for grammatical sentences. These results converge with grammaticality judgment accuracy and response time data in Experiment 2, where weaker number attraction effects obtained for grammatical sentences as compared to ungrammatical ones. Importantly, these asymmetrical effects suggest that they are in fact due to attraction and not to carry-over effects originated while reading the preceding plural attractors, as might have occurred in the self-paced reading task. If the effects shown in ungrammatical sentences were due to carry-over effects, they should also have been detected in grammatical ones.

In sum, behavioral and ERP results from Experiment 2 showed that antecedent-clitic dependencies are also subject to attraction effects and that these effects are detected in ungrammatical sentences only. Our ERP results identified frontal negative components as the main electrophysiological indexes of attraction effects.

### On Self-paced Reading vs. ERP Data

The fact that no clear attraction effects obtained in reading times for antecedent-clitic dependencies (either at clitic position or following word regions) suggests that this type of dependencies are resilient to attraction effects, as previously revealed by Dillon et al. (2013) and Parker and Phillips (2017). However, in Experiment 2 these effects were detectable by methods with finer temporal resolution such as ERPs. Off-line grammaticality judgment measures showed attraction effects in both experiments, but asymmetrical attraction effects were



only obtained in Experiment 2. Certain experimental design variables might have contributed to these differences. For instance, in Experiment 1 grammatical and ungrammatical sentences containing plural antecedents were included, so that a grammaticality judgment task could be performed with sentences containing plural clitics. In contrast, in Experiment 2 all sentences with plural clitics were ungrammatical. Although some researchers counterbalanced this by adding as fillers grammatical sentences containing plural controllers (e.g., Franck et al., 2015: in all three Experiments), many self-paced reading and ERP studies do not (Pearlmutter et al., 1999: Experiments 1 and 2; Shen et al., 2013; Tanner et al., 2014: in all three Experiments), or do not report it (Chen et al., 2007; Lago et al., 2015: in all four Experiments; Wagers et al., 2009: Experiments 2, 4, 5, and 6). The fact that in Experiment 2 a plural clitic was a perfectly reliable signal of an ungrammatical sentence might have lessened the capacity of attractors to elicit attraction. Although the confound between grammaticality and the processing of plural clitics might have had some effect, we believe this cannot be the determinant factor behind the results of Experiment 2. This is because we should expect similar electrophysiological response patterns related to grammaticality (or clitic number) effects in sentences containing number matching and mismatching attractors. However, we report different ERP patterns in grammatical vs. ungrammatical sentences: frontal negativity-P600 vs. only P600, respectively. If participants used a plural-clitic equal ungrammatical task-specific strategy, attraction effects might have diminished overall, augmenting the possibility to detect asymmetrical attraction effects. The impact these design differences might have had on the off-line results obtained in both experiments seem to be reflected in our on-line measures too, where ERP data provided finer grained timing effects than the self-paced reading data. Next, we discuss the main implications of these findings for current models of language processing.

## Fitting the Findings with Models of Agreement Processing

The negative components (e.g., LAN/N400) reported above have been generally interpreted as indexing a greater difficulty to integrate the predicted critical word into the previous context (Friederici et al., 1993; Münte et al., 1993; Friederici, 2002; Rossi et al., 2005). In the case of antecedent-clitic dependencies (e.g., *...el paquete para el/los vecino(s) \*los...*), attraction effects modulate these negative components, revealing a greater difficulty to predict/integrate a plural clitic (*\*los*) that disagrees in number with the singular antecedent (*el paquete*) when it was preceded by a singular attractor (*el vecino*), as compared to when it was preceded by a plural attractor (*los vecinos*).

These results can be accounted for under the cue-based retrieval and similarity-based interference accounts of dependency processing (Lewis and Vasishth, 2005; Wagers et al., 2009). In these models, when the number feature of the dependent element (in our case, the clitic) is encountered, a retrieval mechanism searches for a matching element stored in working memory (in this case, the antecedent NP). Accordingly, the grammaticality effect reported here ensues: when a plural clitic is encountered in a context where all possible antecedents

are singular, the predictability and/or integration of the clitic is most difficult, eliciting a frontal negativity. But when a plural clitic is encountered after a plural attractor, the similarity-based interference of the plural attractor with the singular antecedent to be retrieved from memory leads to erroneously interpreting the plural attractor as the antecedent of the clitic, so that no frontal negativity is elicited.

Conversely, the asymmetric effect emerges because, in grammatical sentences, the number mismatching attractor is not retrieved from memory, either because the retrieval mechanism is not deployed or because it only retrieves antecedent candidates that fully match the features of the clitic. In other words, when the number of the antecedent and the dependent clitic match (i.e., in grammatical sentences), it is assumed that the number of the attractor noun is not retrieved, so that no ERP differences ensue for matching and mismatching attractors. In contrast, when the number of the antecedent and the clitic do not match (i.e., in ungrammatical sentences), a reanalysis process ensues. Thus, a larger frontal negativity is expected in sentences containing number matching attractors as compared to mismatching ones, because illusions of grammaticality only occur in sentences where plural attractor NPs match the number of the plural clitic and can be mistakenly retrieved as their antecedents (Wagers et al., 2009; Phillips et al., 2011). Our results fully support these predictions.

The absence of a frontal negativity in sentences containing mismatching plural attractors is an index of the presence of number attraction effects during antecedent-clitic dependency resolution. Additionally, the asymmetric effect revealed by the absence of ERP components indexing attraction in grammatical sentences (in contrast to the negative component elicited in ungrammatical ones), provides compelling evidence in support of cue-based retrieval models as accounts of attraction effects in comprehension (Wagers et al., 2009; Shen et al., 2013). Unfortunately, our data cannot adjudicate between the possibility that encountering a singular clitic out-competes retrieval of plural antecedent candidates so that the attractor is not retrieved, and the possibility that the dependency is correctly processed without the deployment of retrieval mechanisms. Importantly, the present results cannot be accommodated into feature percolation models because they assume attraction effects are driven by an erroneous number representation in the antecedent NP. Hence, they predict equivalent effects in grammatical and ungrammatical sentences (Nicol et al., 1997; Pearlmutter et al., 1999), contrary to our findings (see also, Wagers et al., 2009; Dillon et al., 2013; Lago et al., 2015). Electrophysiological evidence in Experiment 2 revealed a clear asymmetrical effect of attraction. In sum, the evidence here provides strong support for cue-based retrieval models of dependency resolution in language processing, and are incompatible with alternative feature percolation accounts.

Finally, in addition to the absence of frontal negative components as an electrophysiological index of attraction effects, we also reported that number violations elicited a P600 component both when sentences contained a matching singular attractor and when they contained a mismatching plural attractor. This in turn reveals that, despite the presence of attraction effects, participants could detect the ungrammaticality of sentences containing number violations. These findings

contrast with those reported in Shen et al. (2013), where attraction effects led to the absence of associated ERP components, and in Kaan (2002) and Tanner et al. (2014), where attraction effects caused a reduction in the amplitude of the ungrammaticality/reanalysis related P600. In the former case, the differences in results likely originate from task differences: the grammaticality judgment task we used required participants to explicitly and consciously check and reanalyze sentences for well-formedness, which would encourage the appearance of the P600 even in sentences where number attraction effects occurred, while the comprehension task used by Shen et al. (2013) did not require participants to pay attention to grammaticality. However, the differences between our results and those of Kaan (2002) and Tanner et al. (2014), where attraction effects reduced the amplitude of the P600 component, are harder to explain based on task differences, given that all studies used a grammaticality judgment task. The main differences with regard to our study involves a smaller ratio of ungrammatical sentences (28.6% in our study vs. 50% in theirs); the type of phenomenon explored, where we studied attraction effects in antecedent-clitic dependencies and they did so in subject-verb agreement. These findings might tentatively be interpreted as evidence that antecedent-clitic dependencies tap into processes different from those involved in agreement.

## What Do These Findings Reveal about the Nature of Clitic Dependencies?

Whether Romance clitics are pronouns or agreement morphemes is under debate. Although our experimental approach does not provide direct evidence supporting either type of syntactic analysis, we believe it offers indirect evidence that can be informative. Some studies compared the magnitude of attraction effects elicited in different types of dependencies (Dillon et al., 2013; Parker and Phillips, 2017) and observed that antecedent-reflexive pronoun dependencies were more resilient to attraction than subject-verb agreement. We suggest that these differences correlate with the distinct nature of these two types of dependencies. Although both types of dependencies rely on similar cue-based retrieval mechanisms, antecedent-pronoun dependencies involve a referential dependency between two nominal arguments and weight structural cues more strongly than morphological ones, precluding the erroneous retrieval of non-licensed antecedent candidates. In contrast, subject-verb agreement is a morphological mechanism used to index the arguments of sentences where morphological cues weigh more than structural ones, making the erroneous retrieval of non-licensed attractors possible.

Although a direct comparison of attraction effects between subject-verb and antecedent-clitic dependencies goes beyond the scope of this study, we argue that our results align better with results previously obtained for antecedent-reflexive pronoun than for subject-verb agreement dependencies. Subject-verb agreement resolution shows consistent attraction effects in self-paced reading studies and these effects have been shown to mainly modulate late positive P600 components in ERP experiments. In contrast, antecedent-clitic dependency

resolution is resilient to attraction effects in self-paced reading (Experiment 1; replicating reflexive pronoun studies: Dillon et al., 2013; Parker and Phillips, 2017) and affected early frontal negative ERP components (Experiment 2). Hence, we tentatively interpret the observed resilience of antecedent-clitic dependencies to attraction effects and the fact that they modulate different electrophysiological components than in subject-verb agreement to indicate that antecedent-clitic and verb agreement dependencies constitute different types of linguistic dependencies. Thus, we interpret our indirect evidence to favor the *Clitics as Pronouns Hypothesis* originally proposed by Kayne (1975), which suggests that Spanish object-clitics are processed as pronominal elements.

Regarding the ERP patterns indexing attraction effects in antecedent-clitic dependencies, we observed that attraction effects lead to the absence of a frontal negativity, while in previous studies subject-verb agreement attraction effects modulated the later positive P600 component (Kaan, 2002; Tanner et al., 2014). Although we reckon that linking the modulation of different ERP components as evidence for different types of processes might be seen as speculative, we tentatively argue the difference in the type of associated components signals the different processes involved in the resolution of pronominal and agreement dependencies: attraction effects revealed by frontal negativities might be related to pronominal processing and active retrieval of the lexical representation of possible antecedents and signals difficulty of syntactic/semantic integration of the full arguments (i.e., difficulty of establishing antecedent-clitic pronominal dependencies) (Barkley et al., 2015). Thus, in ungrammatical sentences, no negative component appeared because the plural attractor NP might have been incorrectly identified as the plural clitic antecedent NP. In contrast, attraction effects revealed at later positivities might be mainly related to purely morphological processes like agreement and signal difficulty to integrate and reanalyze morphological features (e.g., in verb agreement dependencies) (Hagoort et al., 1993).

Certainly, further research making direct comparisons between antecedent-clitic and subject-verb agreement dependencies in relatively similar syntactic contexts (i.e., distance between the two agreeing elements, structural position of the elements, whether they are in their canonical position or not, etc.) will help to better identify the processes underlying both structures. Further research ought to provide a fuller and more systematic picture of the main electrophysiological indexes involved in the resolution of the different types on linguistic dependencies across a wider array of languages.

## CONCLUSION

We provide novel evidence regarding the electrophysiological indexes associated to processing mechanisms underlying attraction effects in the comprehension of antecedent-clitic dependencies. Our results show that antecedent-clitic dependencies can be disrupted by an intervening attractor.

Studying this pattern of disruption, we replicate the grammatical asymmetry of attraction effects observed in subject-verb agreement (Wagers et al., 2009; Shen et al., 2013; Tanner et al., 2014, 2016), which supports cue-based retrieval mechanisms of attraction. Finally, despite being resilient to attraction effects in self-paced reading measures, clitic dependencies show electrophysiological indexes of attraction that involve components different from those commonly found for verb agreement (frontal negativities for clitics and late positivities for agreement). These differences, we speculate, suggest that clitic-pronoun and verb agreement dependencies involve distinct processing routines for their resolution. Further research involving more languages and types of dependencies will undoubtedly contribute to shed more detail in this general picture of dependency-processing in language comprehension.

## AUTHOR CONTRIBUTIONS

MS and AZ performed research and analyzed data. MS, AZ, KE, and IL designed, discussed and interpreted findings and wrote the paper.

## REFERENCES

- Acuña-Fariña, J. C., Meseguer, E., and Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua* 143, 108–128. doi: 10.1016/j.lingua.2014.01.013
- Anagnostopoulou, E. (2003). *The Syntax of Ditransitives: Evidence from Clitics*. Berlin: Mouton de Gruyter.
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511801686
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Badecker, W., and Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: evidence from gender and case in Slovak. *J. Mem. Lang.* 56, 65–85. doi: 10.1016/j.jml.2006.08.004
- Barber, H., and Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *J. Cogn. Neurosci.* 17, 137–153. doi: 10.1162/0898929052880101
- Barkley, C., Kluender, R., and Kutas, M. (2015). Referential processing in the human brain: an event-related potential (ERP) study. *Brain Res.* 1629, 143–159. doi: 10.1016/j.brainres.2015.09.017
- Bock, J. K., and Eberhard, K. M. (1993). Meaning, sound, and syntax in English number agreement. *Lang. Cogn. Process.* 8, 57–99. doi: 10.1080/01690969308406949
- Bock, J. K., Eberhard, K. M., and Cutting, J. C. (2004). Producing number agreement: how pronouns equal verbs. *J. Mem. Lang.* 51, 251–278. doi: 10.1016/j.jml.2004.04.005
- Bock, J. K., and Middleton, E. (2011). Reaching agreement. *Nat. Lang. Linguist. Theory* 29, 1033–1069. doi: 10.1007/s11049-011-9148-y
- Bock, K., and Cutting, C. (1992). Regulating mental energy: performance units in language production. *J. Mem. Lang.* 31, 99–127. doi: 10.1016/0749-596X(92)90007-K
- Bock, K., and Miller, C. (1991). Broken agreement. *Cogn. Psychol.* 23, 35–43. doi: 10.1016/0010-0285(91)90003-7
- Bock, K., Nicol, J., and Cutting, J. C. (1999). The ties that bind: creating number agreement in speech. *J. Mem. Lang.* 40, 330–346. doi: 10.1006/jmla.1998.2616
- Borer, H. (1984). *Parametric Syntax: Case Studies in Semitic and Romance Languages*. Dordrecht: Foris Publications. doi: 10.1515/9783110808506

## FUNDING

This research has been supported by grants from the Spanish Ministerio de Economía y Competitividad and Ministerio de Ciencia e Innovación (FFI2014-55733-P; FFI2015-64183-P; RYC-2010-06520, RYC-2013-14722), and the Basque Government (IT665-13).

## ACKNOWLEDGMENTS

We thank the reviewers for the very helpful comments and suggestions provided and for their great contribution to the improvement of this manuscript. We also thank Anna Hatzidaki for her very helpful comments, and Idoia Ros for her help testing participants.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01470/full#supplementary-material>

- Bornkessel, I., and Schlesewsky, M. (2006). The Extended Argument Dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychol. Rev.* 113, 787–821. doi: 10.1037/0033-295X.113.4.787
- Chen, L., Shu, H., Liu, Y., Zhao, J., and Li, P. (2007). ERP signatures of subject-verb agreement in L2 learning. *Biling. Lang. Cogn.* 10, 161–174. doi: 10.1017/S136672890700291X
- Corbett, G. (2006). *Agreement*. Cambridge: Cambridge University Press.
- Coulson, S., King, J., and Kutas, M. (1998). Expect the unexpected: event-related brain response to morphosyntactic violations. *Lang. Cogn. Process.* 13, 21–58. doi: 10.1080/016909698386582
- De Vincenzi, M., Job, R., Di Matteo, R., Angrilli, A., Penolozzi, B., Ciccarelli, L., et al. (2003). Differences in the perception and time course of syntactic and semantic violations. *Brain Lang.* 85, 280–296. doi: 10.1016/S0093-934X(03)00055-5
- Deutsch, A., and Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: evidence from ERPs and eye movements. *J. Mem. Lang.* 45, 200–224. doi: 10.1006/jmla.2000.2768
- Díaz, B., Sebastián-Gallés, N., Erdocia, K., Mueller, J. L., and Laka, I. (2011). On the cross-linguistic validity of electrophysiological correlates of morphosyntactic processing: a study of case and agreement violations in Basque. *J. Neurolinguist.* 24, 357–373. doi: 10.1016/j.jneuroling.2010.12.003
- Dillon, B., Chow, W. Y., Wagers, M., Guo, T., Liu, F., and Phillips, C. (2014). The structure-sensitivity of memory access: evidence from Mandarin Chinese. *Front. Psychol.* 5:1025. doi: 10.3389/fpsyg.2014.01025
- Dillon, B., Chow, W. Y., and Xiang, M. (2016). The relationship between anaphor features and antecedent retrieval: comparing Mandarin ziji and ta-ziji. *Front. Psychol.* 6:1966. doi: 10.3389/fpsyg.2015.01966
- Dillon, B., Mishler, A., Slogett, S., and Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: experimental and modeling evidence. *J. Mem. Lang.* 69, 85–103. doi: 10.1016/j.jml.2013.04.003
- Eberhard, K. M. (1997). The marked effect of number on subject agreement. *J. Mem. Lang.* 36, 147–164. doi: 10.1006/jmla.1996.2484
- Fernández Soriano, O. (1989). Strong pronouns in null subject languages and the Avoid Pronoun Principle. *MIT Work. Pap. Linguist.* 11, 228–240.
- Foucart, A., and Frenck-Mestre, C. (2011). Grammatical gender processing in L2: electrophysiological evidence of the effect of L1-L2 syntactic similarity. *Biling. Lang. Cogn.* 14, 379–399. doi: 10.1017/S136672891000012X

- Foucart, A., and Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *J. Mem. Lang.* 66, 226–248. doi: 10.1016/j.jml.2011.07.007
- Franck, J. (2011). Reaching agreement as a core syntactic process. Commentary of Bock & Middleton 'Reaching Agreement'. *Nat. Lang. Linguist. Theory* 29, 1071–1086. doi: 10.1007/s11049-011-9153-1
- Franck, J., Colonna, S., and Rizzi, L. (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Front. Psychol.* 6:349. doi: 10.3389/fpsyg.2015.00349
- Franco, J. (2000). "Agreement as a continuum: the case of Spanish pronominal clitics," in *Clitic Phenomena in European Languages*, eds F. Beukema and M. den Dikken (Amsterdam: John Benjamins), 147–190. doi: 10.1075/la.30.07fra
- Frenck-Mestre, C., McLaughlin, J., Osterhout, L., and Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: evidence from ERPs. *Acta Psychol.* 128, 528–536. doi: 10.1016/j.actpsy.2007.12.007
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84. doi: 10.1016/S1364-6613(00)01839-8
- Friederici, A. D., Pfeifer, E., and Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cogn. Brain Res.* 1, 183–192. doi: 10.1016/0926-6410(93)90026-2
- Gunter, T. C., Friederici, A. D., and Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J. Cogn. Neurosci.* 12, 556–568. doi: 10.1162/0899892900562336
- Hagoort, P., Brown, C. M., and Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Lang. Cogn. Process.* 8, 439–483. doi: 10.1080/01690969308407585
- Hammer, A., Jansma, B., Lamers, M., and Münte, T. F. (2005). Pronominal reference in sentences about persons or things: an electrophysiological approach. *J. Cogn. Neurosci.* 17, 227–239. doi: 10.1162/08998929053124947
- Hammer, A., Jansma, B. M., Lamers, M., and Münte, T. (2008). Interplay of meaning, syntax and working memory during pronoun resolution investigated by ERPs. *Brain Res.* 1230, 177–191. doi: 10.1016/j.brainres.2008.07.004
- Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007
- Jaeggli, O. (1982). *Topics in Romance Syntax*. Dordrecht: Foris Publications.
- Jaeggli, O. (1986). "Three issues in the theory of clitics: case, doubled NPs, and extraction," in *Syntax and Semantics 19: The Syntax of Pronominal Clitics*, ed. H. Borer (New York, NY: Academic Press), 15–42.
- Jäger, L. A., Engelmann, F., and Vasishth, S. (2015). Retrieval interference in reflexive processing: experimental evidence from Mandarin, and computational modeling. *Front. Psychol.* 6:617. doi: 10.3389/fpsyg.2015.00617
- Jäger, L. A., Engelmann, F., and Vasishth, S. (2017). Similarity-based interference in sentence comprehension: literature review and Bayesian meta-analysis. *J. Mem. Lang.* 94, 316–339. doi: 10.1016/j.jml.2017.01.004
- Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject-verb dependencies: an ERP study. *J. Psycholinguist. Res.* 31, 165–193. doi: 10.1023/A:1014978917769
- Kayne, S. R. (1975). *French Syntax: The Transformational Cycle*. Cambridge, MA: MIT Press.
- Kutas, M., and Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Mem. Cogn.* 11, 539–550. doi: 10.3758/BF03196991
- Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., and Phillips, C. (2015). Agreement attraction in Spanish comprehension. *J. Mem. Lang.* 82, 133–149. doi: 10.1016/j.jml.2015.02.002
- Lamers, M. J. A., Jansma, B. M., Hammer, A., and Münte, T. F. (2006). Neural correlates of semantic and syntactic processes in the comprehension of case marked pronouns: evidence from German and Dutch. *BMC Neurosci.* 7:23. doi: 10.1186/1471-2202-7-23
- Lamers, M. J. A., Jansma, B. M., Hammer, A., and Münte, T. F. (2008). Differences in the processing of anaphoric reference between closely related languages: neurophysiological evidence. *BMC Neurosci.* 9:55. doi: 10.1186/1471-2202-9-55
- Lewis, R. L., and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.* 29, 375–419. doi: 10.1207/s15516709cog0000\_25
- Marchis, M., and Alexiadou, A. (2013). The syntax of clitics revisited: two types of clitics. *Lingua* 127, 1–13. doi: 10.1016/j.lingua.2013.01.002
- Martin-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., and Sommer, W. (2006). Semantic prevalence over syntax during sentence processing: a brain potential study of noun-adjective agreement in Spanish. *Brain Res.* 1093, 178–189. doi: 10.1016/j.brainres.2006.03.094
- Molinari, N., Barber, H. A., Caffarra, S., and Carreiras, M. (2015). On the left anterior negativity (LAN): the case of morphosyntactic agreement: a reply to Tanner et al. *Cortex* 66, 156–159. doi: 10.1016/j.cortex.2014.06.009
- Molinari, N., Barber, H. A., and Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex* 47, 908–930. doi: 10.1016/j.cortex.2011.02.019
- Molinari, N., Vespignani, F., and Job, R. (2008). A deeper reanalysis of a superficial feature: an ERP study on agreement violations. *Brain Res.* 1228, 161–176. doi: 10.1016/j.brainres.2008.06.064
- Monachesi, P. (2005). *The Verbal Complex in Romance: A Case Study in Grammatical Interfaces*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199274758.001.0001
- Münte, T. F., Heinze, H. J., and Mangun, G. (1993). Dissociation of brain activity related to syntactic and semantic aspects of language. *J. Cogn. Neurosci.* 5, 335–344. doi: 10.1162/jocn.1993.5.3.335
- Nevins, A., Dillon, B., Malhotra, S., and Phillips, C. (2007). The role of feature number and feature-type in processing Hindi verb agreement violations. *Brain Res.* 1164, 81–94. doi: 10.1016/j.brainres.2007.05.058
- Nicol, J., and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *J. Psycholinguist. Res.* 18, 5–19. doi: 10.1007/BF01069043
- Nicol, J. L., Forster, K. I., and Veres, C. (1997). Subject-verb agreement processes in comprehension. *J. Mem. Lang.* 36, 569–587. doi: 10.1006/jmla.1996.2497
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Osterhout, L., Bersick, M., and McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Mem. Cogn.* 25, 273–285. doi: 10.3758/BF03211283
- Osterhout, L., McKinnon, R., Bersick, M., and Corey, V. (1996). On the language-specificity of the brain response to syntactic anomalies: is the syntactic positive shift a member of the P300 family? *J. Cogn. Neurosci.* 8, 507–526. doi: 10.1162/jocn.1996.8.6.507
- Osterhout, L., and Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *J. Mem. Lang.* 34, 739–773. doi: 10.1006/jmla.1995.1033
- Parker, D., and Phillips, C. (2017). Reflexive attraction in comprehension is selective. *J. Mem. Lang.* 94, 272–290. doi: 10.1016/j.jml.2017.01.002
- Patil, U., Vasishth, S., and Lewis, R. L. (2016). Retrieval interference in syntactic processing: the case of reflexive binding in English. *Front. Psychol.* 7:329. doi: 10.3389/fpsyg.2016.00329
- Pearlmutter, N. J. (2000). Linear versus hierarchical agreement feature processing in comprehension. *J. Psycholinguist. Res.* 29, 89–98. doi: 10.1023/A:1005128624716
- Pearlmutter, N. J., Garnsey, S. M., and Bock, K. (1999). Agreement processes in sentence comprehension. *J. Mem. Lang.* 41, 427–456. doi: 10.1006/jmla.1999.2653
- Phillips, C., Wagers, M., and Lau, E. F. (2011). "Grammatical illusions and selective fallibility in real-time language comprehension," in *Experiments at the Intersection, Syntax and Semantics*, Vol. 37, ed. J. Runner (Bingley: Emerald Publications), 147–180. doi: 10.1108/S0092-456320110000037009
- R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rivas, A. M. (1977). *A Theory of Clitics*. Ph.D. thesis Massachusetts Institute of Technology, Cambridge, MA.
- Rohde, D. (2001). *Linger*. Available at: <http://tedlab.mit.edu/~dr/Linger/> [accessed January 20, 2017].
- Rossi, E., Kroll, J. F., and Dussias, P. E. (2014). Clitic pronouns reveal the time course of processing gender and number in a second language. *Neuropsychologia* 62, 11–25. doi: 10.1016/j.neuropsychologia.2014.07.002

- Rossi, S., Gugler, M. F., Hahne, A., and Friederici, A. D. (2005). When word category information encounters morphosyntax: an ERP study. *Neurosci. Lett.* 384, 228–233. doi: 10.1016/j.neulet.2005.04.077
- Schmitt, B. M., Lamers, M., and Münte, T. F. (2002). Electrophysiological estimates of biological and syntactic gender violation during pronoun processing. *Cogn. Brain Res.* 14, 333–346. doi: 10.1016/S0926-6410(02)00136-2
- Severens, E., Jansma, B. M., and Hartsuiker, R. J. (2008). Morphophonological influences on the comprehension of subject-verb agreement: an ERP study. *Brain Res.* 1228, 135–144. doi: 10.1016/j.brainres.2008.05.092
- Shen, E. Y., Staub, A., and Sanders, L. D. (2013). Event-related brain potential evidence that local nouns affect subject-verb agreement processing. *Lang. Cogn. Process.* 28, 498–524. doi: 10.1080/01690965.2011.650900
- Silva-Pereyra, J., and Carreiras, M. (2007). An ERP study of agreement features in Spanish. *Brain Res.* 1185, 201–211. doi: 10.1016/j.brainres.2007.09.029
- Silva-Pereyra, J., Gutierrez-Sigut, E., and Carreiras, M. (2012). An ERP study of coreference in Spanish: semantic and grammatical gender cues. *Psychophysiology* 49, 1401–1411. doi: 10.1111/j.1469-8986.2012.01446.x
- Sportiche, D. (1998). *Partitions and Atoms of Clause Structure: Subjects, Agreement, Case and Clitics*. New York, NY: Routledge.
- Strozer, J. R. (1976). *Clitics in Spanish*. Ph.D. thesis, University of California, Los Angeles, Los Angeles, CA.
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *J. Mem. Lang.* 48, 542–562. doi: 10.1016/S0749-596X(02)00536-3
- Suñer, M. (1988). The role of agreement in clitic-doubled constructions. *Nat. Lang. Linguist. Theory* 6, 391–434. doi: 10.1007/BF00133904
- Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: a commentary on “Grammatical agreement processing in reading: ERP findings and future directions” by Molinaro et al., 2014. *Cortex* 66, 149–155. doi: 10.1016/j.jml.2014.07.003
- Tanner, D., Grey, S., and Van Hell, J. G. (2016). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology* 54, 248–259. doi: 10.1111/psyp.12788
- Tanner, D., Nicol, J., and Brehm, L. (2014). The time course of feature interference in agreement comprehension: multiple mechanisms and asymmetrical attraction. *J. Mem. Lang.* 76, 195–215. doi: 10.1016/j.jml.2014.07.003
- Torrego, E. (1988). *A DP Analysis of Spanish Nominals*. Boston, MA: University of Massachusetts.
- Uriagereka, J. (1995). Aspects of the syntax of clitic placement in western romance. *Linguist. Inq.* 26, 79–124.
- Wagers, M. W., Lau, E. F., and Phillips, C. (2009). Agreement attraction in comprehension: representations and processes. *J. Mem. Lang.* 61, 206–237. doi: 10.1016/j.jml.2009.04.002
- Xu, X., Jiang, X., and Zhou, X. (2013). Processing biological gender and number information during Chinese pronoun resolution: ERP evidence for functional differentiation. *Brain Cogn.* 81, 2223–2236. doi: 10.1016/j.bandc.2012.11.002
- Zawiszewski, A., and Friederici, A. D. (2009). Processing object-verb agreement in canonical and non-canonical word orders in Basque: evidence from event-related brain potentials. *Brain Res.* 1284, 161–179. doi: 10.1016/j.brainres.2009.05.099
- Zawiszewski, A., Gutiérrez, E., Fernández, B., and Laka, I. (2011). Language distance and non-native syntactic processing: evidence from event-related potentials. *Biling. Lang. Cogn.* 14, 400–411. doi: 10.1017/S1366728910000350
- Zwicky, A. M. (1977). *On Clitics*. Bloomington, IN: Indiana University Linguistics Club.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Santesteban, Zawiszewski, Erdocia and Laka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Toward Cognitively Constrained Models of Language Processing: A Review

Margreet Vogelzang<sup>1,2\*</sup>, Anne C. Mills<sup>1,3</sup>, David Reitter<sup>4</sup>, Jacolien Van Rij<sup>1</sup>, Petra Hendriks<sup>1</sup> and Hedderik Van Rijn<sup>2,5</sup>

<sup>1</sup>Center for Language and Cognition Groningen, University of Groningen, Groningen, Netherlands, <sup>2</sup>Department of Experimental Psychology, University of Groningen, Groningen, Netherlands, <sup>3</sup>Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom, <sup>4</sup>College of Information Sciences and Technology, The Pennsylvania State University, State College, PA, United States, <sup>5</sup>Department of Statistical Methods and Psychometrics, University of Groningen, Groningen, Netherlands

## OPEN ACCESS

### Edited by:

Ángel J. Gallego,  
Universitat Autònoma  
de Barcelona, Spain

### Reviewed by:

Mireille Besson,  
Cognitive Neuroscience Laboratory,  
CNRS and Aix-Marseille University,  
France  
Cristiano Chesì,  
Istituto Universitario di Studi Superiori  
di Pavia (IUSS), Italy

### \*Correspondence:

Margreet Vogelzang  
margreetvogelzang@gmail.com

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Communication

**Received:** 03 April 2017

**Accepted:** 23 August 2017

**Published:** 08 September 2017

### Citation:

Vogelzang M, Mills AC, Reitter D,  
Van Rij J, Hendriks P and Van Rijn H  
(2017) Toward Cognitively  
Constrained Models of Language  
Processing: A Review.  
Front. Commun. 2:11.  
doi: 10.3389/fcomm.2017.00011

Language processing is not an isolated capacity, but is embedded in other aspects of our cognition. However, it is still largely unexplored to what extent and how language processing interacts with general cognitive resources. This question can be investigated with cognitively constrained computational models, which simulate the cognitive processes involved in language processing. The theoretical claims implemented in cognitive models interact with general architectural constraints such as memory limitations. This way, it generates new predictions that can be tested in experiments, thus generating new data that can give rise to new theoretical insights. This theory-model-experiment cycle is a promising method for investigating aspects of language processing that are difficult to investigate with more traditional experimental techniques. This review specifically examines the language processing models of Lewis and Vasishth (2005), Reitter et al. (2011), and Van Rij et al. (2010), all implemented in the cognitive architecture Adaptive Control of Thought—Rational (Anderson et al., 2004). These models are all limited by the assumptions about cognitive capacities provided by the cognitive architecture, but use different linguistic approaches. Because of this, their comparison provides insight into the extent to which assumptions about general cognitive resources influence concretely implemented models of linguistic competence. For example, the sheer speed and accuracy of human language processing is a current challenge in the field of cognitive modeling, as it does not seem to adhere to the same memory and processing capacities that have been found in other cognitive processes. Architecture-based cognitive models of language processing may be able to make explicit which language-specific resources are needed to acquire and process natural language. The review sheds light on cognitively constrained models of language processing from two angles: we discuss (1) whether currently adopted cognitive assumptions meet the requirements for language processing, and (2) how validated cognitive architectures can constrain linguistically motivated models, which, all other things being equal, will increase the cognitive plausibility of these models. Overall, the evaluation of cognitively constrained models of language processing will allow for a better understanding of the relation between data, linguistic theory, cognitive assumptions, and explanation.

**Keywords:** language processing, sentence processing, linguistic theory, cognitive modeling, Adaptive Control of Thought—Rational, cognitive resources, computational simulations

## INTRODUCTION

Language is one of the most remarkable capacities of the human mind. Arguably, language is not an isolated capacity of the mind but is embedded in other aspects of cognition. This can be seen in, for example, linguistic recursion. Although linguistic recursion (e.g., “the sister of the father of the cousin of...”) could in principle be applied infinitely many times, if the construction becomes too complex we will lose track of its meaning due to memory constraints (Gibson, 2000; Fedorenko et al., 2013). Even though there are ample examples of cognitive resources like memory playing a role in language processing (e.g., King and Just, 1991; Christiansen and Chater, 2016; Huettig and Janse, 2016), it is still largely unexplored to what extent language processing and general cognitive resources interact. That is, which general cognitive resources and which language processing-specific resources are used for language processing? For example, is language processing supported by the same memory system that is used in other cognitive processes? In this review, we will investigate to what extent general cognitive resources limit and influence models of linguistic competence. To this end, we will review cognitively constrained computational models of language processing implemented in the cognitive architecture Adaptive Control of Thought—Rational (ACT-R) and evaluate how general cognitive limitations influence linguistic processing in these models. These computational cognitive models explicitly implement theoretical claims, for example about language, based on empirical observations or experimental data. The evaluation of these models will generate new insights about the interplay between language and other aspects of cognition.

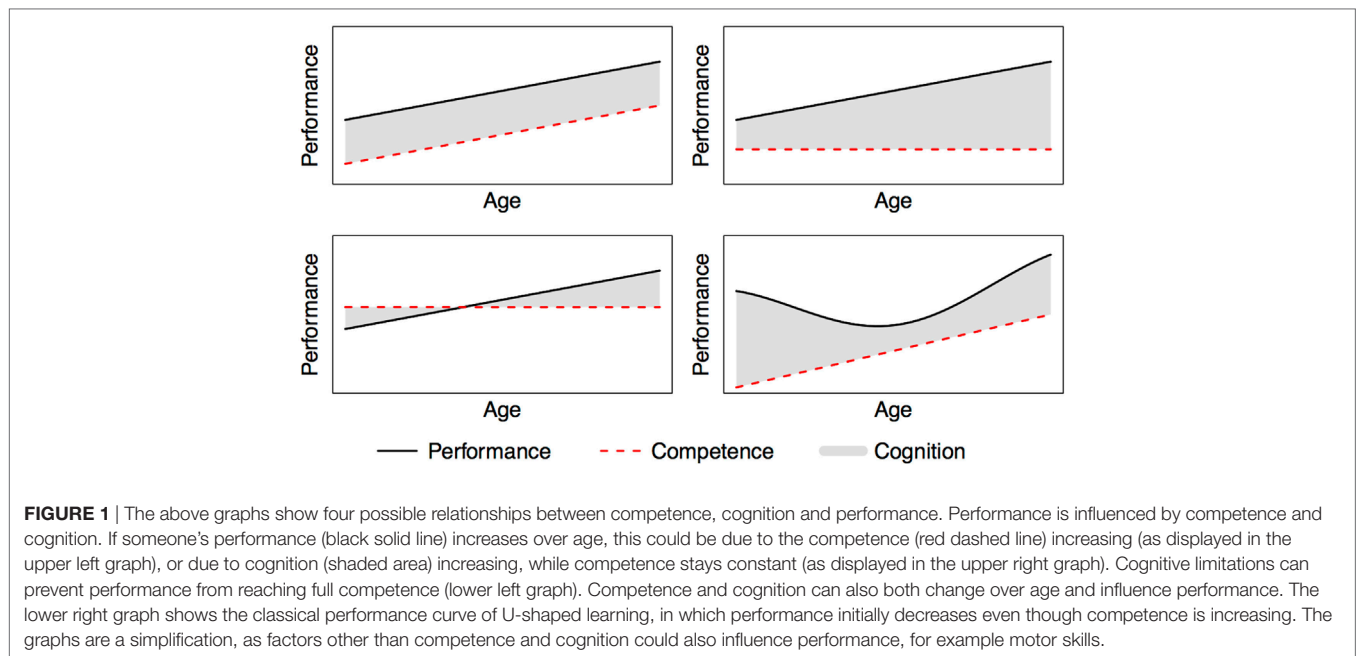
Memory is one of the most important general cognitive principles for language processing. In sentence processing, words have to be processed rapidly, because otherwise the memory of the preceding context, necessary for understanding the complete sentence, will be lost (Christiansen and Chater, 2016). Evidence that language processing shares a memory system with other cognitive processes can be found in the relation between general working memory tests and linguistic tests. For example, individual differences in working memory capacity have been found to play a role in syntactic processing (King and Just, 1991), predictive language processing (Huettig and Janse, 2016), and discourse production (Kuijper et al., 2015). Besides memory, other factors like attentional focus (Lewis et al., 2006) and processing speed (Hendriks et al., 2007) have been argued to influence linguistic performance. Thus, it seems apparent that language processing is not an isolated capacity but is embedded in other aspects of cognition. This claim conflicts with the traditional view that language is a specialized faculty (cf. Chomsky, 1980; Fodor, 1983). It is therefore important to note that computational cognitive models can be used to investigate both viewpoints, i.e., to investigate to what extent general cognitive resources can be used in language processing but also to investigate to what extent language is a specialized process. It has also been argued that language processing is a specialized process that is nevertheless influenced by a range of general cognitive resources (cf. Newell, 1990; Lewis, 1996). Therefore, we argue that the potential influence and limitations

of general cognitive resources should be taken into account when studying theories of language processing.

To be able to account for the processing limitations imposed by a scarcity of cognitive resources, theories of language need to be specified as explicitly as possible with regards to, for example, processing steps, the incrementality of processing, memory retrievals, and representations. This allows for a specification of what belongs to linguistic competence and what belongs to linguistic performance (Chomsky, 1965): competence is the knowledge a language user has, whereas performance is the output that a language user produces, which results from his competence in combination with other (cognitive) factors (see **Figure 1** for examples). Many linguistic theories have been argued to be theories of linguistic competence that abstract away from details of linguistic performance (Fromkin, 2000). These theories rarely make explicit how the step from competence to performance is made. In order to create a distinction between competence and performance, an increasing emphasis is placed on grounding linguistic theories empirically by creating the step from an abstract theory to concrete, testable predictions (cf. e.g., Kempen and Hoenkamp, 1987; Roelofs, 1992; Baayen et al., 1997; Reitter et al., 2011). Formalizing language processing theories explicitly thus means that the distinction between linguistic competence and linguistic performance can be explained and makes it possible to examine which cognitive resources, according to a language processing theory, are needed to process language (see also Hale, 2011).

The importance of explicitly specified linguistic theories that distinguish between competence and performance can be seen in the acquisition of verbs. Children show a U-shaped learning curve (see Pauls et al., 2013 for an overview, U-shaped learning curve is depicted in **Figure 1**) when learning past tenses of verbs, using the correct irregular form first (e.g., the past tense *ate* for *eat*), then using the incorrect regular form of irregular verbs (e.g., *eated*), before using the correct irregular form again. It is conceivable that whereas children’s performance initially decreases, children are in the process of learning how to correctly form irregular past tenses and therefore have increasing competence (cf. Taatgen and Anderson, 2002). In this example, explicitly specifying the processing that is needed to form verb tenses and how this processing uses general cognitive resources could explain why children’s performance does not match their competence. Another example of performance deviating from competence can be seen in the comprehension and production of pronouns: whereas 6-year-old children generally produce pronouns correctly (they have the competence, see Spender et al., 2009), they often make mistakes in pronoun interpretation (they show reduced performance, Chien and Wexler, 1990).

Especially when different linguistic theories have been put forward to explain similar phenomena, it is important to be able to compare and test the theories on the basis of concrete predictions. Linguistic theories are often postulated without considering cognitive resources. Therefore, it is important to investigate how well these theories perform under realistic cognitive constraints; this will provide information about their cognitive plausibility. Cognitively constrained computational models (from now on: cognitive models) are a useful tool to compare linguistic theories



while taking into account the limitations imposed by a scarcity of cognitive resources and can be used to investigate the relation between underlying linguistic competence and explicit predictions about performance. Thus, by implementing a linguistic theory into a cognitive model, language processing is embedded in other aspects of cognition, and the extent can be investigated to which assumptions about general cognitive resources influence models of linguistic competence.

As cognitive models, we will consider computational models simulating human processing that are constrained by realistic and validated assumptions about human processing. Such cognitive models can generate new predictions that can be tested in further experiments, generating new data that can give rise to new implementations. This theory-model-experiment cycle is a promising method for investigating aspects of language processing that are difficult to investigate with standard experimental techniques, which usually provide insight into performance (e.g., behavior, responses, response times), but not competence. Cognitive models require linguistic theories, that usually describe competence, to be explicitly specified. This way, the performance of competing linguistic theories, which often have different approaches to the structure and interpretation of language, can be investigated using cognitive models. Contrary to other computational modeling methods, cognitive models simulate the processing of a single individual. Because of this, it can be investigated how individual variations in cognitive resources (which can be manipulated in a model) influence a linguistic theory's performance.

The comparison of cognitive models that use different linguistic approaches is most straightforward when they make use of the same assumptions about cognitive resources, and thus are implemented in the same cognitive architecture. This review will therefore focus on cognitive models developed in the same domain-general cognitive architecture, ACT-R (Anderson et al., 2004). There are several other cognitive architectures available

(e.g., EPIC: Kieras and Meyer, 1997; NENGO: Stewart et al., 2009), but in order to keep the assumptions about general cognitive resources roughly constant, this review will only consider models implemented in ACT-R. Over the past years, several linguistic phenomena have been implemented in ACT-R, such as metaphors (Budiu and Anderson, 2002), agrammatism (Stocco and Crescentini, 2005), pronominal binding (Hendriks et al., 2007), and presupposition resolution (Brasoveanu and Dotlačil, 2015). In order to obtain a broad view of cognitively constrained models of linguistic theories, we will examine three models of different linguistic modalities (comprehension, production, perspective taking), that all take a different linguistic approach, in depth: the syntactic processing model of Lewis and Vasishth (2005), the syntactic priming model of Reitter et al. (2011), and the pronoun processing model of Van Rij et al. (2010). By examining models of different linguistic modalities that take different linguistic approaches, we aim to provide a more unified understanding of how language processing is embedded within general cognition, and investigate how proficient language use is achieved. The selected models are all bounded by the same assumptions about cognitive capacities and seriality of processing as provided by the cognitive architecture ACT-R, which makes them optimally comparable. Their comparison will provide insight into the extent to which assumptions about general cognitive resources influence models of linguistic competence.

This paper is organized as follows. First, we will discuss the components of ACT-R that are most relevant in our discussion of language processing models, in order to explain how cognitive resources play a role in this architecture. Then, we will outline the different linguistic approaches that are used in the models. Finally, we will discuss the selected ACT-R models of language processing in more detail. Importantly, it will be examined how general cognitive resources are used in the models and how these cognitive resources and linguistic principles interact.



## BASIC ACT-R COMPONENTS

Adaptive Control of Thought—Rational (Anderson, 1993, 2007; Anderson et al., 2004) is a cognitive architecture in which models can be implemented to simulate a certain process or collection of processes. Of specific interest for this review is the simulation of language-related processes, such as interpreting or producing a sentence. Cognitive models in ACT-R are restricted by general cognitive resources and constraints embedded in the ACT-R architecture. Examples of such cognitive resources, that are of importance when modeling language, are memory, processing speed, and attention. By implementing a model of a linguistic theory in ACT-R, one can thus examine how this linguistic theory behaves in interaction with other aspects of cognition.

Adaptive Control of Thought—Rational aims to explain human cognition as the interaction between a set of functional modules. Each module has a specific function, such as perception, action, memory, and executive function [see Anderson et al. (2004) for an overview]. Modules can be accessed by the model through buffers. The information in these buffers represents information that is in the focus of attention. Only the information that is in a buffer can be readily used by the model. An overview of the standard ACT-R modules and buffers is shown in **Figure 2**. The modules most relevant for language processing, the declarative memory module and the procedural memory module, will be discussed in more detail below.

The declarative memory stores factual information as chunks. Chunks are pieces of knowledge that can store multiple properties, such as that there is a cat with the *name* “Coco,” whose *color* is “gray.” The information in a chunk can only be used after the chunk has been retrieved from the declarative memory and has been placed in the corresponding retrieval buffer. In order to retrieve information from memory, a retrieval request must be made. Only chunks with an activation that exceeds a

predetermined activation threshold can be retrieved. The higher the activation of a chunk, the more likely it is to be retrieved. The base-level activation of a chunk increases when a chunk is retrieved from memory, but decays over time. This way, the recency and frequency of a chunk influence a chunk’s activation, and thereby its chance of recall and its retrieval time (in line with experimental findings, e.g., Deese and Kaufman, 1957; Allen and Hulme, 2006). Additionally, information that is currently in the focus of attention (i.e., in a buffer) can increase the probability that associated chunks are recalled by adding spreading activation to a chunk’s base-level activation. The activation of chunks can additionally be influenced by noise, occasionally causing a chunk with less activation to be retrieved over a chunk with more activation.

Whereas the declarative memory represents factual knowledge, the procedural memory represents knowledge about how to perform actions. The procedural memory consists of production rules, which have an if-then structure. An example of the basic structure of a production rule is as follows:

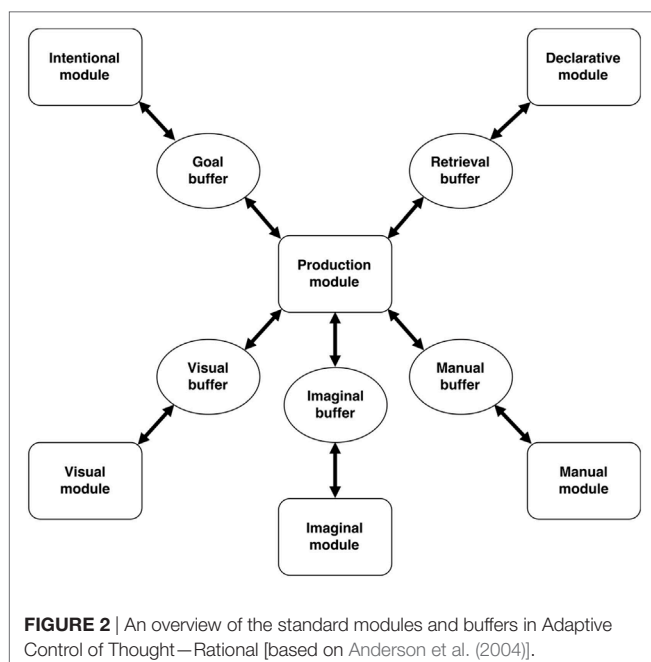
```

IF
  a new word is attended
THEN
  retrieve lexical information about this word from memory
  
```

The THEN-part of a production rule is executed when the IF-part matches the current buffer contents. Production rules are executed one by one. If the conditions of several production rules are met, the one with the highest utility is selected. This utility reflects the usefulness the rule has had in the past and can be used to learn from feedback, both positively and negatively (for more detail on utilities, see Anderson et al., 2004). New production rules can be learned on the basis of existing rules and declarative knowledge (production compilation, Taatgen and Anderson, 2002).

Several general cognitive resources and further resources that are important for language processing are incorporated in the ACT-R architecture, such as memory, speed of processing, and attention. Long-term memory corresponds to the declarative module in ACT-R. Short-term or working memory is not incorporated as a separate component in ACT-R (Borst et al., 2010) but emanates from the interaction between the buffers and the declarative memory. Daily et al. (2001) proposed that the function of working memory can be simulated in ACT-R by associating relevant information with information that is currently in focus (through spreading activation). Thus, working memory capacity can change as a result of a change in the amount of spreading activation in a model.

Crucially, all above mentioned operations take time. Processing in ACT-R is serial, meaning that only one retrieval from declarative memory and only one production rule execution can be done at any point in time (this is known as the serial processing bottleneck, see Anderson, 2007). The retrieval of information from declarative memory is faster and more likely to succeed if a chunk has a high activation (for details see Anderson et al., 2004). Because a chunk’s activation increases when it is retrieved, chunks that have been retrieved often will have a high activation and will therefore be retrieved more quickly. Production rules in



ACT-R take a standard amount of time to fire (50 ms). Rules that are often used in succession can merge into a new production rule. These new rules are a combination of the old rules that were previously fired in sequence, making the model more efficient. Thus, increasing activation and production compilation allow a model's processing speed to increase through practice and experience.

As described, memory and processing speed are examples of general cognitive principles in ACT-R, that will be important when implementing models that perform language processing. In the next section, three linguistic approaches will be discussed. These approaches are relevant for the three cognitive models reviewed in the remainder of the paper.

## LINGUISTIC APPROACHES

Cognitive models can be used to implement any linguistic approach, and as such are not bound to one method or theory. In principle any of the theories that have been proposed in linguistics to account for a speaker's linguistic competence, such as Combinatorial Categorical Grammar (Steedman, 1988), construction grammar (Fillmore et al., 1988), generative syntax (Chomsky, 1970), Head-driven Phrase Structure Grammar (Pollard and Sag, 1994), Lexical Functional Grammar (Bresnan, 2001), Optimality Theory (OT) (Prince and Smolensky, 1993), Tree-Adjoining Grammar (Joshi et al., 1975), and usage-based grammar (Bybee and Beckner, 2009) could be implemented in a cognitive model. Note that this does not imply that any linguistic theory or approach can be implemented in any cognitive model, as cognitive models place restrictions on what can and cannot be modeled. Different linguistic approaches tend to entertain different assumptions, for example about what linguistic knowledge looks like (universal principles, violable constraints, structured lexical categories, grammatical constructions), the relation between linguistic forms and their meanings, and the levels of representation needed. This then determines whether and how a particular linguistic approach can be implemented in a particular cognitive model.

In this review, we will discuss three specific linguistic approaches that have been implemented in cognitive models, which allows us to compare how general cognitive resources influence the implementation and output (e.g., responses, response times) of these modeled linguistic approaches. The three linguistic approaches that will be discussed have several features in common but also differ in a number of features: X-bar theory (Chomsky, 1970), Combinatorial Categorical Grammar (Steedman, 1988), and OT (Prince and Smolensky, 1993). These linguistic approaches are implemented in the cognitive models discussed in the next section.

Generative syntax uses X-bar theory to build syntactic structures (Chomsky, 1970). X-bar theory reflects the assumption that the syntactic representation of a clause is hierarchical and can be presented as a binary branching tree. Phrases are built up around a head, which is the principal category. For example, the head of a verb phrase is the verb, and the head of a prepositional phrase is a preposition. To the left or right of this head, other phrases can be attached in the hierarchical structure.

Combinatorial Categorical Grammar (CCG) (Steedman, 1988) builds the syntactic structure of a sentence in tandem with the representation of the meaning of the sentence. It is a strongly lexicalized grammar formalism, that proceeds from the assumption that the properties of the grammar follow from the properties of the words in the sentence. That is, each word has a particular lexical category that specifies how that word can combine with other words, and what the resulting meaning will be. In addition, CCG is surface-driven and reflects the assumption that language is processed and interpreted directly, without appealing to an underlying—invisible—level of representation. For one sentence, CCG can produce multiple representations (Steedman, 1988; Reitter et al., 2011). This allows CCG to build syntactic representations incrementally, from left to right.

The linguistic framework of OT (Prince and Smolensky, 1993) reflects the assumption that language is processed based on constraints on possible outputs (words, sentences, meanings). Based on an input, a set of output candidates is generated. Subsequently, these potential outputs are evaluated based on hierarchically ranked constraints; stronger constraints have priority over weaker constraints. The optimal output is the candidate that satisfies the set of constraints best. The optimal output may be a form (in language production) or a meaning (in language comprehension).

## Commonalities and Differences

X-bar theory, CCG, and OT have different assumptions about how language is structured. X-bar theory builds a syntactic structure, whereas CCG builds both a syntactic and a semantic representation, and OT builds either a syntactic representation (in language production) or a semantic representation (in language comprehension). Nevertheless, these theories can all be used for the implementation of cognitive models of language processing. In the next section, three cognitive models of language processing will be discussed in detail, with a focus on how the linguistic approaches are implemented and how they interact with other aspects of cognition.

## COGNITIVE MODELS OF LANGUAGE PROCESSING

In the following sections, three cognitive language models will be described: the sentence processing model of Lewis and Vasishth (2005), the syntactic priming model of Reitter et al. (2011), and the pronoun processing model of Van Rij et al. (2010). The model of Lewis and Vasishth (2005) uses a parsing strategy that is based on X-bar theory, the model of Reitter et al. (2011) uses CCG, and the model of Van Rij et al. (2010) uses OT. The models will be evaluated based on their predictions of novel empirical outcomes and how they achieve these predictions (for example how many parameters are fitted, cf. Roberts and Pashler, 2000). After describing the models separately, the commonalities and differences between these models will be discussed. Based on this, we will review how the interaction between general cognitive resources in ACT-R and linguistic principles from specific linguistic theories can be fruitful in studying cognitive assumptions of linguistic theories.

## Modeling Sentence Processing as Skilled Memory Retrieval

The first model that we discuss is the sentence processing model of Lewis and Vasishth (2005). This model is a seminal model forming the basis for many later language processing models (a.o., Salvucci and Taatgen, 2008; Engelmann et al., 2013; Jäger et al., 2015). Lewis and Vasishth's (Lewis and Vasishth, 2005) sentenced processing model (henceforth the L&V model) performs syntactic parsing based on memory principles: when processing a complete sentence, maintaining the part of the sentence that is already processed in order to integrate it with new incoming information requires (working) memory. The aim of the L&V model is to investigate how working memory processes play a role in sentence processing.

### Theoretical Approach

The L&V model uses left-corner parsing (Aho and Ullman, 1972), based on X-bar theory (Chomsky, 1970), to build a syntactic representation of the sentence. The left corner (LC) parser builds a syntactic structure of the input sentence incrementally, and predicts the upcoming syntactic structure as new words are encountered. Thus, LC parsing uses information from the words in the sentence to predict what the syntactic structure of that sentence will be. In doing this, LC parsing combines top-down processing, based on syntactic rules, and bottom-up processing, based on the words in a sentence. An example sentence is (1).

(1) The dog ran.

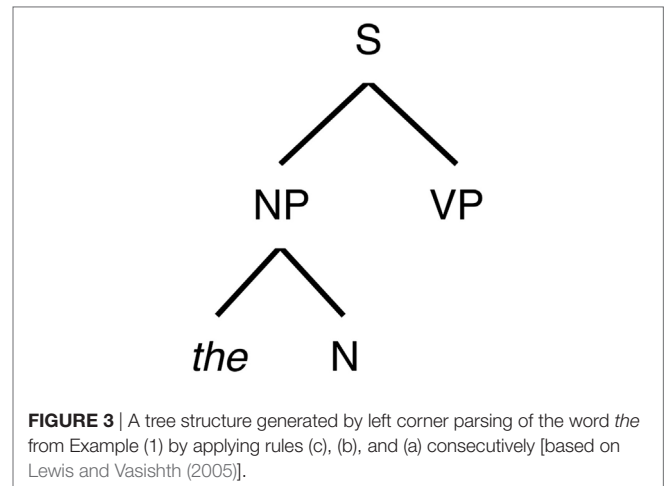
Left corner parsing is based on structural rules, such as those given below as (a)–(d). These structural rules for example state that a sentence can be made up of a noun phrase (NP) and a verb phrase [rule (a)], and that a NP can be made up of a determiner and a noun [rule (b)]. An input (word) is nested under the lefthand-side (generally an overarching category) of a structural rule if that rule contains the input on its LC. For example, in sentence (1), *the* is a determiner (Det) according to structural rule (c), which itself is on the LC of rule (b) and thus it is nested under an NP. This NP is on the LC of rule (a). The result of applying these rules is the phrase-structure tree shown in **Figure 3**.

- (a) S → NP VP
- (b) NP → Det N
- (c) Det → the
- (d) N → dog

Importantly, the generated tree also contains syntactic categories that have not been encountered yet (like N and VP in **Figure 3**), so it contains a prediction of the upcoming sentence structure. When the next word, *dog*, is now encountered, it can be integrated with the existing tree immediately after applying rule (d).

### Implementation

The L&V model parses a sentence on the basis of guided memory retrievals. Declarative memory is used as the short- and long-term memory needed for sentence processing. The declarative memory holds lexical information as well as any syntactic



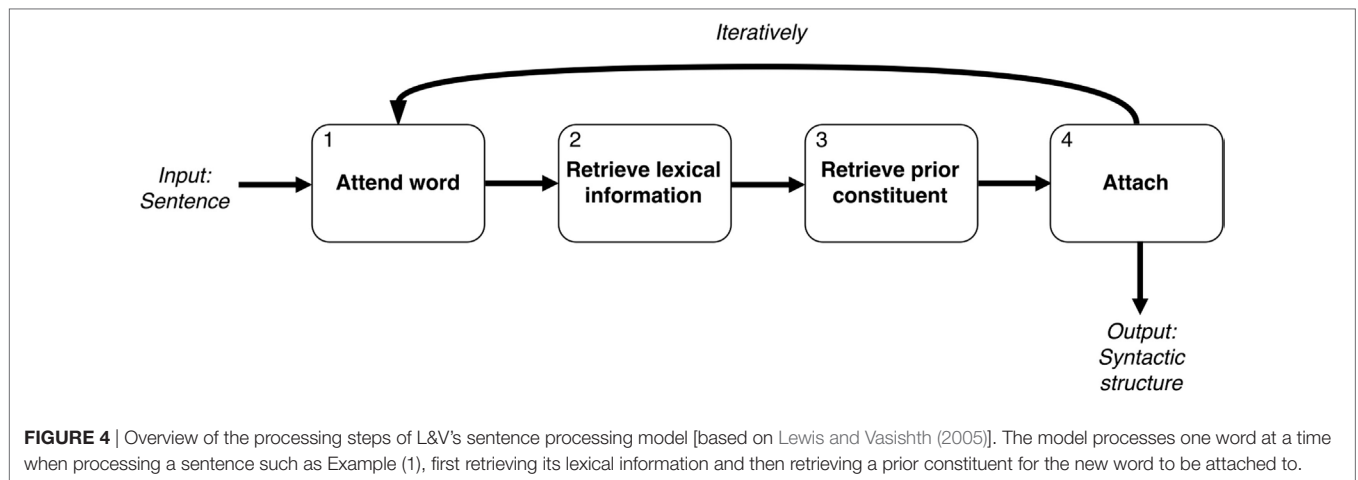
structures that are built during sentence processing. The activation of these chunks is influenced by the standard ACT-R declarative memory functions, and so their activation (and with this their retrieval probability and latency) is influenced by the recency and frequency with which they were used. Similarity-based interference occurs because the effectiveness of a retrieval request is reduced as the number of items associated with the specific request increases.

Grammatical knowledge however is not stored in the declarative memory but is implemented as procedural knowledge in production rules. That is, the knowledge about how sentences are parsed is stored in a large number of production rules, which interact with the declarative memory when retrieving lexical information or constituents (syntactic structures).

The L&V model processes a sentence word for word using the LC parsing algorithm described in Section “Theoretical Approach.” An overview of the model's processing steps is shown in **Figure 4**. After a word is attended [for example, *the* from Example (1), Box 1], lexical information about this word is retrieved from memory and stored in the lexical buffer (Box 2). Based on the syntactic category of the word and the current state of the model, the model looks for a prior constituent that the new syntactic category could be attached to (Box 3). In our example, *the* is a determiner and it is the first word, so a syntactic structure with a determiner will be retrieved. The model then creates a new syntactic structure by attaching the new word to the retrieved constituent (Box 4). A new word is then attended [*dog* in Example (1), Box 1]. This cycle continues until no new words are left to attend.

### Evaluation

Lewis and Vasishth (2005) presented several simulation studies, showing that their model can account for reading times from experiments. The model also accounts for the effects of the length of a sentence (short sentences are read faster than long sentences) and structural interference (high interference creates a bigger delay in reading times than low interference) on unambiguous and garden-path sentences. With a number of additions (that are outside the scope of this review), the model can be made to cope



with gapped structures and embedded structures, as well as local ambiguity (see Lewis and Vasishth, 2005, for more detail).

### Predictions

Lewis and Vasishth (2005) compared their output to existing experiments, rather than making explicit predictions about new experiments. The model does however provide ideas about *why* any discrepancies between the model and the fitted data occur, which could be seen as predictions, although these predictions have not been tested in new experiments. For example, in a simulation comparing the L&V models' simulated reading times of subject relative clauses vs. object relative clauses to data from Grodner and Gibson (2005), the model overestimates the cost of object-gap filling for object relative clauses. The prediction following from the model is that adjusting the latency, a standard ACT-R parameter that influences the time it takes to perform a chunk retrieval, would reduce the difference between model and data. Thus, the prediction is that the retrieval latency of chunks may be lower in this type of language processing than in other cognitive processes.

### Linguistic Principles

X-bar theory is a widely known approach to syntactic structure. Although already previously implemented as an LC parser (Aho and Ullman, 1972), it is interesting to examine this linguistic approach in interaction with memory functions. Importantly, the use of LC parsing allowed the L&V model to use a top-down (prediction-based, cf. Chesi, 2015) as well as bottom-up (input-based, cf. Chomsky, 1993) processing, which increases its efficiency.

### Cognitive Principles

Many of the cognitive principles used in the L&V model are taken directly from ACT-R: memory retrievals are done from declarative memory, the grammatical knowledge needed for parsing is incorporated in production rules, and sentences are processed serially (word by word). Memory plays a very important role in the model, as processing sentences requires memory of the recent past. For all memory functions, the same principles of

declarative memory are used as would be used for non-linguistic processes. For the L&V model, the standard ACT-R architecture was expanded with a lexical buffer, which holds a lexical chunk after it is retrieved from the declarative memory. Thus, the model assumes the use of general memory functions for language processing, but added a specific attention component to store linguistic (lexical) information that is in the focus of attention.

The speed of processing required for language processing is achieved in the L&V model by keeping the model's processing down to the most efficient way to do things: the processing of a word takes a maximum of three production rules and two memory retrievals, serially. This however includes only the syntactic processing, and not, for example, any semantic processing. It remains to be investigated therefore how the model would function if more language processing elements, that take additional time to be executed due to the serial processing bottleneck, are added.

### Limitations and Future Directions

Although the simulations show a decent fit when compared to data from several empirical experiments, there are a number of phenomena for which a discrepancy is found between the simulation data and some of the experimental data. Specifically, the L&V model overestimates effects of the length of a sentence and underestimates interference effects. Lewis and Vasishth (2005) indicated that part of this discrepancy may be resolved by giving more weight to decay and less weight to interference in the model, but leave the mechanisms responsible for length effects and interference effects open for future research.

Lewis and Vasishth (2005) acknowledged that the model is a first step to modeling complete sentence comprehension and indicated that future extensions might lie in the fields of semantic and discourse processing, the interaction between lexical and syntactic processing, and investigating individual performance based on working memory capacity differences. Indeed, this sentence processing model is an influential model that has served as a building block for further research. For example, Engelmann et al. (2013) used the sentence processing model to study the relation between syntactic processing and eye movements, Salvucci and Taatgen (2008) used the model in their research of

multitasking, and Van Rij et al. (2010) and Vogelzang (2017) build their OT model of pronoun resolution on top of L&V's syntactic processing model.

## Modeling Syntactic Priming in Language Production

A second model discussed in this paper is the ACT-R model of Reitter et al. (2011). Their model (henceforth the RK&M model) investigates syntactic priming in language production. Speakers have a choice between different words and grammatical structures to express their ideas. They tend to repeat previously encountered grammatical structures, a pattern of linguistic behavior that is referred to as syntactic or structural priming (for a review, see Pickering and Ferreira, 2008). For example, Bock (1986) found that when speakers were presented with a passive construction such as *The boy was kissed by the girl* as a description of a picture, they were more likely to describe a new picture using a similar syntactic structure. Effects of priming have been detected with a range of syntactic constructions, including NP variants (Cleland and Pickering, 2003), the order of main and auxiliary verbs (Hartsuiker and Westenberg, 2000), and other structures, in a variety of languages (Pickering and Ferreira, 2008), and in children (Huttenlocher et al., 2004; Van Beijsterveldt and Van Hell, 2009), but also syntactic phrase-structure rules in general (Reitter et al., 2006; Reitter and Moore, 2014).

In the literature, a number of factors that interact with priming have been identified:

- **Cumulativity:** priming strengthens with each copy of the primed construction (Jaeger and Snider, 2008).
- **Decay:** the probability of occurrence of a syntactic construction decays over time (Branigan et al., 1999).
- **Lexical boost:** lexically similar materials increase the chance that priming will occur (Pickering and Branigan, 1998).
- **Inverse frequency interaction:** priming by less frequent constructions is stronger (Scheepers, 2003).

Besides these factors, differences have been found between fast, short-term priming and slow, long-term adaptation, which is a learning effect that can persist over several days (Bock et al., 2007; Kaschak et al., 2011b). These two different priming effects have been suggested to use separate underlying mechanisms (Hartsuiker et al., 2008), and as such may rely on different cognitive resources.

Syntactic priming is seen as an important effect by which to validate models of syntactic representations and associated learning. Several other models of syntactic priming were proposed (Chang et al., 2006; Snider, 2008; Malhotra, 2009), but none of these are able to account for all mentioned factors as well as short and long term priming. The goal of the RK&M model is thus to account for all types of syntactic priming within a cognitive architecture.

### Theoretical Approach

The RK&M model is based on a theoretical approach that explains priming as facilitation of lexical-syntactic access. The model bases its syntactic composition process on a broad-coverage grammar framework, CCG (see Linguistic Approaches,

Steedman, 1988, 2000). Categorical Grammars use a small set of combinatory rules and a set of parameters to define the basic operations that yield sentences in a specific language. Most specific information is stored in the lexicon. With the use of CCG, the RK&M model implements the idea of combinatorial categories as in Pickering and Branigan's (Pickering and Branigan, 1998) model.

In CCG, the syntactic process is the result of combinations of adjacent words and phrases (in constituents). Unlike classical phrase-structure trees, however, the categories that classify each constituent reflect its syntactic and semantic status by stating what other components are needed before a sentence results. For example, the phrase *loves toys* needs to be combined with a NP to its left, as in Example 2. This phrase is assigned the category S\NP. Similarly, the phrase *Dogs love* requires a NP to its right to be complete, thus, its category is S//NP. Many analyses (derivations) of a given sentence are possible in CCG.

(2) Dogs love toys.

Combinatory Categorical Grammar allows the RK&M model to generate a syntactic construction incrementally, so that a speaker can start speaking before the entire sentence is planned. However, it also allows the planning of a full sentence before a speaker starts speaking. CCG is generally underspecified and generates more sentences than would be judged acceptable. The RK&M model at least partially addresses this over-generation by employing memory-based ACT-R mechanisms, which also help in providing a cognitively plausible version of a language model.

### Implementation

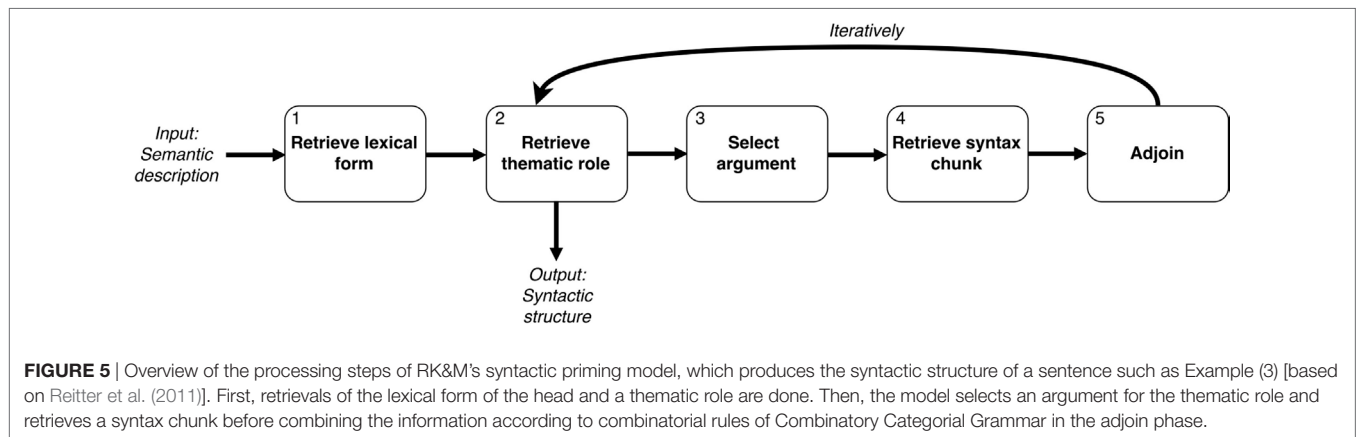
In the RK&M model, lexical forms and syntactic categories are stored in chunks in declarative memory. The activation of any chunk in ACT-R is determined by previous occurrences, which causes previously used, highly active chunks to have a higher retrieval probability, creating a priming effect.

The RK&M model additionally uses spreading activation to activate all syntax chunks that are associated with a lexical form, creating the possibility to express a meaning in multiple ways. Some ways of expressing a meaning are more frequent in language than others, and therefore the amount of spreading activation from a lexical form to a syntax chunk is mediated by the frequency of the syntactic construction. This causes more frequent forms to have a higher activation and therefore to be more likely to be selected. However, a speaker's choice of syntactic construction can vary on the basis of priming and noise.

To make its theoretical commitments to cue-based, frequency- and recency-governed declarative retrieval, as well as its non-commitments to specific production rules and their timing more clear, the RK&M model was implemented first in ACT-R 6, and then in the ACT-UP implementation of the ACT-R theory (Reitter and Lebiere, 2010).

### Syntactic Realization

The RK&M model takes a semantic description of a sentence as input and creates a syntactic structure for this input. The serially executed processing steps of the model are shown in **Figure 5** and will be explained on the basis of Example (3).



### (3) Sharks bite.

First, the model retrieves a lexical form for the head of the sentence (Box 1). In Example (3), this head will be the verb *bite*. Then the most active thematic role is retrieved from memory (Box 2), which would be the “agent-role” in our example. If no next thematic role can be retrieved, the entire sentence has been generated and an output can be given. The model then identifies the argument associated with the retrieved thematic role and retrieves a lexical form for this argument (Box 3). In the case of the agent-role in Example (3), this will be *sharks*. Following, the model retrieves a syntax chunk that is associated with the retrieved lexical form (Box 4). The lexical form was *sharks*, and the corresponding syntax chunk will thus indicate that this is an NP, and that it needs a verb to its right (S/VP). Finally, the model adjoins the new piece of syntactic information with the syntactic structure of the phrase thus far (Box 5), according to the combinatorial rules of CCG. The model then goes back to retrieving the next thematic role (Box 2) and repeats this process until the entire sentence has been generated.

### Priming

Within the language production process, syntactic choice points (Figure 5, Box 4) will occur, during which a speaker decides between several possible syntactic variants. The model needs to explicate the probability distribution over possible decisions at that point. This can be influenced by priming.

The time course of priming is of concern in the RK&M model. Immediately after a prime, repetition probability is strongly elevated. The model uses two default ACT-R mechanisms, base-level learning and spreading activation, to account for long-term adaptation and short-term priming. Short-term priming emerges from a combination of two general memory effects: (1) rapid temporal decay of syntactic information and (2) cue-based memory retrieval subject to interfering and facilitating semantic information (Reitter et al., 2011). Long-term priming effects in the model emerge from the increase in base-level activation that occurs when a chunk is retrieved.

### Evaluation

In the RK&M model, base-level learning and spreading activation account for long-term adaptation and short-term priming,

respectively. By simulating a restricted form of incremental language production, it accounts for (a) the inverse frequency interaction (Scheepers, 2003; Reitter, 2008; Jaeger and Snider, 2013); (b) the absence of a decay in long-term priming (Hartsuiker and Kolk, 1998; Bock and Griffin, 2000; Branigan et al., 2000; Bock et al., 2007); and (c) the cumulativeness of long-term adaptation (Jaeger and Snider, 2008). The RK&M model also explains the lexical boost effect and the fact that it only applies to short-term priming, because semantic information is held in short-term memory and serves as a source of activation for associated syntactic material.

The model uses lexical-syntactic associations as in the residual-activation account (Pickering and Branigan, 1998). However, learning remains an implicit process, and routinization (acquisition of highly trained sequences of actions) may still occur, as it would in implicit learning accounts.

The RK&M model accounts for a range of priming effects, but despite providing an account of grammatical encoding, it has not been implemented to explain how speakers construct complex sentences using the broad range of syntactic constructions found in a corpus.

### Predictions

Because semantic information is held in short-term memory and serves as a source of activation for associated syntactic material, the RK&M model predicts that lexical boost occurs with the repetition of any lexical material with semantic content, rather than just with repeated head words. This prediction was confirmed with corpus data (Reitter et al., 2011) and also experimentally (Scheepers et al., 2017). The RK&M model also predicts that only content words cause a lexical boost effect. This prediction was not tested on the corpus, although it is compatible with prior experimental results using content words (Corley and Scheepers, 2002; Schoonbaert et al., 2007; Kootstra et al., 2012) and semantically related words (Cleland and Pickering, 2003), and the insensitivity of priming to closed-class words (Bock and Kroch, 1989; Pickering and Branigan, 1998; Ferreira, 2003).

The model predicted cumulativeness of prepositional-object construction priming, and it suggested that double-object constructions are ineffective as primes to the point where cumulativeness cannot be detected. In an experimental study published

later by another lab (Kaschak et al., 2011a), this turned out to be the case.

### Linguistic Principles

An important aspect of the RK&M model is that it uses CCG. This allows the model to realize syntactic constructions both incrementally and non-incrementally, without storing large amounts of information. CCG can produce multiple representations of the input at the same time, which reflect the choices that a speaker can make. CCG has enjoyed substantial use on large-scale problems in computational linguistics in recent years. Still, how much does this theoretical commitment (of CCG) limit the model's applicability? The RK&M model relies, for its account of grammatical encoding, on the principles of incremental planning made possible by categorial grammars. However, for its account of syntactic priming, the deciding principle is that the grammar is lexicalized, and that syntactic decisions involve lower-frequency constructions that are retrieved from declarative (lexical) memory. Of course, ACT-R as a cognitive framework imposes demands on what the grammatical encoder can and cannot do, chiefly in terms of working memory: large, complex symbolic representations such as those necessary to process subtrees in Tree-Adjoining Grammar (Joshi et al., 1975), or large feature structures of unification-based formalisms such as Head-driven Phrase Structure Grammar (Pollard and Sag, 1994) would be implausible under the assumptions of ACT-R.

### Cognitive Principles

The RK&M model's linguistic principles are intertwined with cognitive principles in order to explain priming effects. Declarative memory retrievals and the accompanying activation boost cause frequently used constructions to be preferred. Additionally, the model uses the default ACT-R component of spreading activation to give additional activation to certain syntax chunks, increasing the likelihood that a specific syntactic structure will be used. Working memory capacity is not specified in the RK&M model.

The RK&M model is silent with respect to the implementation of its grammatical encoding algorithms. Standard ACT-R provides for production rules that represent routinized skills. These rules are executed at a rate of one every 50 ms. Whether that is fast enough for grammatical encoding when assuming a serial processing bottleneck, and how production compilation can account for fast processing, is unclear at this time. Production compilation, in ACT-R, can combine a sequence of rule invocations and declarative retrievals into a single, large and efficient production rule. An alternative explanation may be that the production rule system associated with the syntactic process is not implemented by the *basal ganglia*, the brain structure normally associated with ACT-R's production rules, but by a language-specific region such as *Broca's area*. This language-specific region may allow for faster processing.

### Limitations and Future Directions

Some effects related to syntactic priming remain unexplained by the RK&M model. For example, the repetition of thematic and semantic assignments between sentences (Chang et al., 2003) is not a consequence of retrieval of lexical-syntactic material. A

future ACT-R model can make use of working memory accounts (cf. Van Rij et al., 2013) to explain repetition preferences leading to such effects.

## Modeling the Acquisition of Object Pronouns

The third and final model that is discussed, is Van Rij et al.'s (2010) model for the acquisition of the interpretation of object pronouns (henceforth the RR&H model). In languages such as English and Dutch, an object pronoun (*him* in Example 4) cannot refer to the local subject (*the penguin* in Example 4, cf. e.g., Chomsky, 1981). Instead, it must refer to another referent in the context, in our example *the sheep*. In contrast, reflexives such as "zichzelf" (*himself, herself*) can only refer to the local subject.

(4) Look, a penguin and a sheep. The penguin is hitting him/himself.

Children up to age seven allow the unacceptable interpretation of the object pronoun "him" (*the penguin*), although children perform adult-like on the interpretation of reflexives from the age of four (e.g., Chien and Wexler, 1990; Philip and Coopmans, 1996). Interestingly, children as early as 4 years old show adult-like production of object pronouns and reflexives (e.g., De Villiers et al., 2006; Spenader et al., 2009). The ACT-R model is used to investigate why children show difficulties interpreting object pronouns, but not interpreting reflexives or producing object pronouns or reflexives.

### Theoretical Account

To explain the described findings on the interpretation of object pronouns and reflexives, Hendriks and Spenader (2006) proposed that children do not lack the linguistic knowledge needed for object pronoun interpretation but fail to take into account the speaker's perspective. According to this account, formulated within OT (Prince and Smolensky, 1993, see Linguistic Approaches), object pronouns compete with reflexives in their use and interpretation.

In the account of Hendriks and Spenader (2006), two grammatical constraints guide the production and interpretation of pronouns and reflexives. "Principle A" is the strongest constraint, which states that reflexives have the same reference as the subject of the clause. In production, Hendriks and Spenader assume a general preference for producing reflexives over pronouns, which is formulated in the constraint "Avoid Pronouns."

Hendriks and Spenader (2006) argue that the interpretation of object pronouns is not ambiguous for adults, because they take into account the speakers' perspective: if the speaker wanted to refer to the subject (e.g., *the penguin* in Example 4), then the speaker would have used a reflexive in accordance with the constraint Principle A. When the speaker did not use a reflexive, therefore, an adult listener should be able to conclude that the speaker must have wanted to refer to another referent. Although this account can explain the asymmetry in children's production and interpretation of object pronouns, it does not provide a theory on how children acquire the interpretation of object pronouns. To investigate this question, the theoretical account of Hendriks and Spenader was implemented in ACT-R (Van Rij et al., 2010; see also Hendriks et al., 2007).

### Implementation

An overview of the RR&H model is presented in **Figure 6**. The process of finding the optimal meaning for a form (in comprehension) or finding the optimal form for a meaning (in production) was implemented in ACT-R as a serial process. To illustrate the process, consider the interpretation of the pronoun *him*.

#### Using Grammatical Constraints

When interpreting a pronoun, two consecutive production rules request the retrieval of two candidate interpretations from the model's declarative memory (Box 1 and Box 2 in **Figure 6**). The two candidate interpretations are the co-referential interpretation (i.e., reference to the referent expressed by the local subject, e.g., *the penguin* in Example 4) and the disjoint interpretation (i.e., reference to another referent in the discourse, such as *the sheep* in Example 4). Consequently, a production rule requests the retrieval of a grammatical constraint from declarative memory. The chunk that represents the constraint Principle A has the highest activation because it is the strongest constraint and is retrieved from memory first (see Box 3).

On the basis of the retrieved constraint, the two candidate interpretations are evaluated (Box 4 and 5). If one of the candidates violates the constraint, the RR&H model tries to replace that candidate by a new candidate (Box 4 and Box 2). If it cannot find a new candidate in memory, the remaining candidate is selected as the optimal interpretation.

If the input was a pronoun, however, none of the candidate interpretations violates Principle A. Therefore, both candidate interpretations are still possible (Box 5). In this situation, the

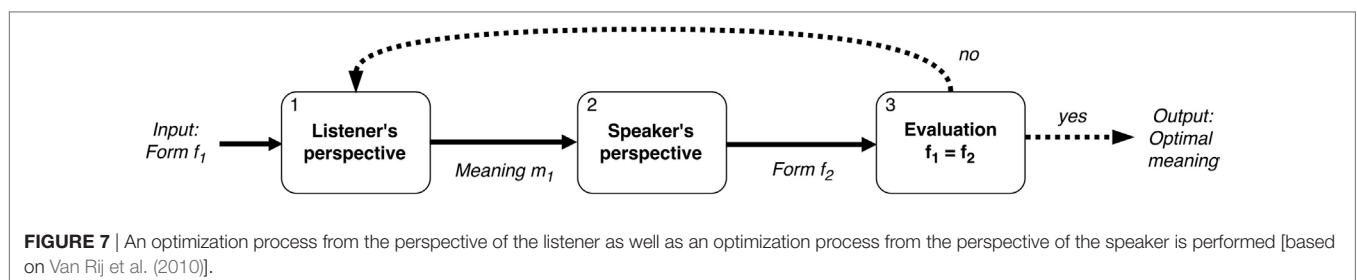
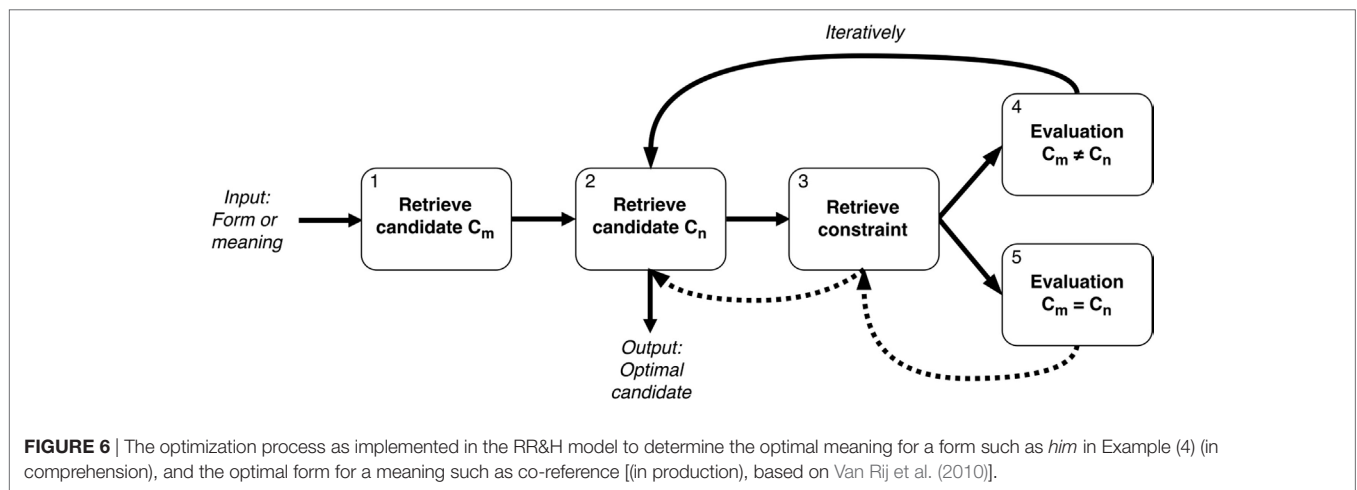
RR&H model retrieves a new constraint (Box 3), Avoid Pronouns. This constraint cannot distinguish between the two candidate meanings either, because it only applies to forms. As both the co-referential and the disjoint interpretation are still possible, the model randomly selects one of the two candidates as the optimal interpretation. The random choice between two optimal candidates reflects children's behavior in the interpretation of object pronouns.

#### Perspective Taking

After selecting the optimal interpretation, the RR&H model takes the speaker's perspective to verify whether the speaker indeed intended to express the selected interpretation (see **Figure 7**). Taking the speaker's perspective, the model uses the same optimization mechanism, but now the input is the *meaning* (optimal interpretation) selected in the previous step when taking the listener's perspective ( $m_1$ ), and the output is the optimal form to express that meaning ( $f_2$ ).

Continuing with the example of processing an object pronoun, the model could have selected the co-referential interpretation as the interpretation of the object pronoun when taking the listener's perspective. In that situation, the input ( $m_1$ ) for the second optimization step, using the speaker's perspective, would be the co-referential interpretation. The output of the second optimization step ( $f_2$ ) is the reflexive form, because the constraint Avoid Pronouns favors the use of a reflexive over a pronoun.

After the two optimization steps, a new production rule fires that compares the initial input (the object pronoun) with the output (a reflexive, **Figure 7** Box 3). As these forms are not identical





in our example, the model concludes that a co-referential interpretation is not intended by the speaker: the speaker would have used a reflexive rather than a pronoun to express a co-referential interpretation. As a consequence, the model will take an alternative candidate interpretation, the disjoint interpretation, and will check if the speaker could have intended a disjoint interpretation.

Alternatively, if the model had selected a disjoint interpretation for a pronoun during the first optimization step, the input for the speaker's perspective ( $m_1$ ) would be a disjoint interpretation. The constraint Principle A would cause the model to select a pronoun rather than a reflexive for expressing the disjoint interpretation ( $f_2$ ). As the original input ( $f_1$ , a pronoun) and the output ( $f_2$ , also a pronoun) are identical, the model concludes that the speaker indeed intended a disjoint interpretation.

Although children are expected to use the same perspective taking mechanism as adults, it is assumed that children's processing is initially too slow to complete this process. The time for pronoun resolution is limited: When the next word comes, the model stops processing the pronoun and redirects its attention to the new word. Gradually however, children's processing becomes more efficient due to ACT-R's default mechanism of production compilation (Taatgen and Anderson, 2002). This way, the process becomes more efficient, and over time it is possible to take the perspective of the speaker into account in interpretation.

## Evaluation

The RR&H model explains the delay in object pronoun acquisition as arising from the interaction between general cognitive principles and specific linguistic constraints. The model simulations show that children's non-adult-like performance does not necessarily arise from differences in linguistic knowledge or differences in processing mechanism but may arise because children lack processing efficiency.

## Predictions

From the RR&H model simulations, a new prediction was formulated: when children receive sufficient time for pronoun interpretation, they will show more adult-like performance on object pronoun interpretation. Van Rij et al. (2010) tested this prediction by slowing down the speech rate. They found that children indeed performed significantly more adult-like on object pronoun interpretation when they were presented with slowed-down speech compared to normal speech. A second prediction of the RR&H model is that the use of perspective taking in pronoun interpretation is dependent on the input frequency of pronouns. With higher input frequency, the process becomes more efficient in a shorter time (Van Rij et al., 2010; Hendriks, 2014).

## Linguistic Principles

The linguistic principles incorporated in the RR&H model is rooted in OT. The underlying idea in OT is that an in principle infinite set of potential candidates is evaluated on the basis of all constraints of the grammar. The serial optimization mechanism implemented in the model is a more constrained version of optimization: the two most likely candidates are compared using the constraints that are most relevant in the context. In this respect, the optimization mechanism could be applied to other

linguistic (and non-linguistic) phenomena and is thus potentially generalizable.

## Cognitive Principles

Several general cognitive principles are used in the RR&H model. Production compilation learning allowed the model to gradually derive an efficient variant of the general cognitive skill of perspective taking that is specialized for object pronoun interpretation. This specialization mechanism has been applied to model other linguistic and non-linguistic phenomena (e.g., Taatgen and Anderson, 2002). Through the increased efficiency of production rules, as well as through increasing activation of candidates and constraints that were used for pronoun interpretation, the model's processing speed increases over time.

The RR&H model uses ACT-R's declarative memory for the storage and retrieval of candidates and constraints. However, no discourse processing was included in the model, and no working memory component was used. Therefore, a remaining question is whether, contrary to what is assumed in other research (Christiansen and Chater, 2016), processing speed limitations on pronoun processing are not imposed by working memory limitations, but by processing efficiency limitations (cf. Kuijper, 2016).

RR&H's account of the difference between children's and adults' processing of pronouns crucially follows from the serial processing bottleneck assumption, as it assumes that children have the knowledge necessary to use bidirectional optimization, including all relevant linguistic knowledge, but cannot make use of it due to time limitations. Proceduralization is used as the explanation for how children arrive at adult performance given the serial processing bottleneck.

## Limitations and Future Directions

A potential limitation of RR&H's object pronoun processing model is that it is not yet clear how to determine the two most likely candidates or how the model can decide what the most relevant constraint is. Another simplification is that both candidate referents were introduced in the previous sentence. An interesting extension of the model would be one in which the discourse status of the referents would also be taken into account (cf. Van Rij et al., 2013). The extended model would need to integrate factors such as first-mention, frequency, recency, grammatical role and role parallelism (Lappin and Leass, 1994), and semantic role (Kong et al., 2009) to account for topicality and the discourse prominence of referents (Grosz et al., 1995), which plays an important role in pronoun resolution (Spencer et al., 2009).

Another future direction for this research would be to investigate why children as early as 4 years old in languages such as Italian and Spanish do not allow unacceptable reference to the local subject for object pronouns (Italian: McKee, 1992; for an overview on Italian see Belletti and Guasti, 2015; Spanish: Baauw, 2002), in contrast to children in languages such as English and Dutch. Thus, this cognitive model could be applied to investigate cross-linguistic variation.

## Commonalities and Differences

In the previous sections, we discussed three language processing models in ACT-R that were based on different linguistic

approaches. The models were all implemented in the same cognitive architecture, so they are all constrained by the same limitations on cognitive resources. This allows for their comparison, which can provide information about how different aspects of language processing interact with non-linguistic aspects of cognition, and how models addressing different linguistic phenomena can be integrated. In this section, we will discuss the commonalities and differences between these models in more detail, so it can be examined to which extent assumptions about general cognitive resources influence implementations of these specific linguistic approaches. Additionally, their comparison will provide an overview of some choices that can be made when implementing a language processing model, such as how to represent (grammatical) knowledge, and how these choices can directly impact how cognitive resources influence the model. The models' main differences lie in (1) the language modality, (2) the linguistic approach they take, and (3) how grammatical knowledge is represented.

As for the different language modalities investigated in the three models, the model of Lewis and Vasishth (2005) focuses on sentence interpretation and builds the syntactic representations needed for interpretation. In contrast, the model of Reitter et al. (2011) focuses on sentence production. The model of Van Rij et al. (2010) again focuses on sentence interpretation but includes a sentence production component in its implementation of perspective taking. So, the selected models show that cognitive models can perform both sentence processing as needed for interpretation and sentence processing as needed for production. As the selected models are merely example implementations of linguistic approaches, this shows how versatile cognitive modeling can be.

A second difference between the three models is that the models all take a different linguistic approach, as Lewis and Vasishth (2005) used LC parsing based on X-bar theory, Reitter et al. (2011) used CCG, and Van Rij et al. (2010) used OT. Although a working cognitive model does not prove the necessity of a particular linguistic approach, it shows its sufficiency: the model of Lewis and Vasishth (2005), for example, shows that LC parsing is sufficient to account for experimental data on sentence processing. It should be noted that the three linguistic approaches need not be mutually exclusive. For example, it is conceivable that a model processes sentences based on LC parsing and uses OT to interpret ambiguous pronouns (cf. Van Rij, 2012; Vogelzang, 2017). Additionally, it should be noted that all three theories have been treated as approaches that have remained unquestioned, whereas variations of these approaches may be worth while to consider (cf., e.g., Osborne et al., 2011).

A final important difference between the models is how grammatical knowledge is represented. In Lewis and Vasishth's (Lewis and Vasishth, 2005) model, lexical information and syntactic structures are stored in declarative memory, but grammatical rules are incorporated as procedural knowledge in production rules. Therefore, their grammatical rules are not subject to the activation functions associated with the declarative memory but are subject to the time constraints of production rule execution. This is different from the model of Reitter et al. (2011), which stores lexical forms as well as syntactic categories as chunks in

the declarative memory, and therefore also incorporates the grammatical rules in the declarative memory. The model of Van Rij et al. (2010) incorporates grammatical rules as chunks in the declarative memory. So, the models incorporate grammatical knowledge in different ways, which has consequences for the influence of general cognitive resources on grammatical knowledge. Specifically, knowledge stored in declarative memory is subject to ACT-R's principles concerning memory activation and retrieval time, whereas knowledge stored in procedural memory is subject to ACT-R's principles concerning production rule execution time.

Although the three models differ in several respects, they also have a number of important features in common. The most important ones that we will discuss are (1) the restrictions placed on the model performance by general cognitive resources, (2) the assumption of a serial processing bottleneck, and (3) the generation of quantitative predictions.

As all models were implemented in ACT-R, the performance of all models is constrained by the same restrictions on cognitive resources. So, although the models focus on different linguistic phenomena and use different representations, they all use, for example, the same functions of declarative memory for the activation of chunks. Furthermore, they all use the same distinction between procedural and declarative memory and incorporate the constraint that information can only be actively used by the model once it is retrieved from declarative memory. Using the same cognitive architecture therefore makes these different models comparable with regard to how the representations are influenced by cognitive resources.

Another constraint within all the models, also imposed by the cognitive architecture, is the serial processing bottleneck (Anderson, 2007). In ACT-R, only one production rule execution or memory retrieval can be performed at a time. Using serial processing increases the time it takes to perform multiple processing steps. Therefore, the serial processing bottleneck creates timing constraints for the models, influencing predictions about performance. We will discuss the implications of this serial processing bottleneck in more detail in the Section "Discussion."

Finally, the last commonality is that all models can generate quantitative predictions. In general, linguistic theories only discuss competence and do not address performance and do not explain why the observed performance may not match the competence. Thus, linguistic theories do not explain, for example, why speakers may use a certain form in 80% of the cases, but a different form in the other cases. By implementing theoretical approaches in cognitive models, quantitative predictions about why performance does not match competence can be generated.

## DISCUSSION

In this review, we investigated to what extent general cognitive resources influence concretely implemented models of linguistic competence. To this end, we examined the language processing models of Lewis and Vasishth (2005), Reitter et al. (2011), and Van Rij et al. (2010). In this section, we will discuss the benefits and limitations of using a cognitive architecture to implement and investigate theories of linguistic competence, and to what

extent general cognitive resources influence performance on the basis of these theories.

Cognitive architectures provide a framework for implementing theories of linguistic competence in a validated account of general cognitive resources related to learning and memory. The three specific models that we discussed showed that the cognitive architecture ACT-R on the one hand provides sufficient freedom to implement different linguistic theories in a plausible manner, and on the other hand sufficiently constrains these theories to account for several differences between linguistic competence and performance. Implementing a linguistic theory in a cognitive architecture forces one to specify, among other things, assumptions about how lexical, syntactic, and semantic knowledge is represented and processed in our mind. These specifications are necessarily constrained by general cognitive resources. Therefore, general cognitive resources such as memory and processing speed also constrain performance on the basis of linguistic theories and are crucial for investigating this performance in a cognitively plausible framework.

By implementing a theory of linguistic competence in a cognitive model, it can be evaluated whether a linguistic theory can account for experimental performance data. The distinction between competence and performance is an advantage of cognitive models over abstract linguistic theories (reflecting competence) and standard experimental measures (measuring performance). A cognitive model thus can not only be used to model performance but can also be used to investigate the reason why full competence may not be reached (e.g., because of memory retrieval limitations: Van Maanen and Van Rijn, 2010, processing speed limitations: Van Rij et al., 2010, or the use of an incorrect strategy: Arslan et al., 2017). As such, cognitive models can account for patterns of linguistic performance that were traditionally accounted for by positing a separate parsing module in the mind specifically for language processing (e.g., Kimball, 1973; Frazier and Fodor, 1978). This line of argumentation has also been explored by Hale (2011), who argues that linguistic theories need to be specified not just on Marr's computational level, but that it is necessary to specify theories at a level of detail so that they can be implemented, step-by-step, in an algorithmic-level framework and yield precise predictions about behavior. The comparison of models described in this review makes explicit which assumptions have to be made in the cognitive model to incorporate particular linguistic theories. All three cognitive models discussed in this review have been applied to fit human data. In many of these cases, the model could account for the general trends in the data, if not the complete data set. As such, all three models provided an explicit relation between data, theory, and explanation. Although not all models made novel predictions that could be tested in new experiments, this is a strength of cognitive modeling and therefore something every paper on cognitive modeling should include. Adding novel predictions shows that (1) the model was not just fitted to existing data and (2) the model is falsifiable. The latter is important, because falsifiable models allow a theory to be disproven. Providing novel predictions allows other researchers to test these, and gather either support for or evidence against a specific theory.

An additional benefit of cognitive modeling is that individual differences can be investigated. By manipulating, for example, the amount of experience (Van Rij et al., 2010), the amount of working memory capacity (Van Rij et al., 2013), or the rate of forgetting in memory (Sense et al., 2016), different performance levels can be achieved. This way, different individuals can be modeled and it can be investigated why certain mistakes may be made (explanations could be, for example limited experience, limited memory capacity, limited attention span). By combining different simulated individuals, group effects may be explained (Van Rij et al., 2010).

There are, however, also some limitations to modeling language processing in a cognitive architecture. First, all three models that were discussed can account for specific linguistic phenomena, but these only form a small part of language. Scalability is an issue for many models, as expanding their coverage and making them more complex (for example, by combining a model that performs full semantic processing with a model that performs full syntactic processing) will make models slower in any architecture that assumes serial processing. Specifically, although the model of Van Rij et al. (2010) uses the serial processing bottleneck explicitly to account for children's performance errors, both Lewis and Vasishth (2005) and Reitter et al. (2011) suggest that their models may struggle with this assumption when expanded. It is thus important to keep in mind that the discussed, relatively small, serially implemented models of language processing were sufficient to fit to experimental data, but the serial processing bottleneck may prove to be too strict for sentence processing when a complete language processing model is developed. Moreover, the discussed models are abstractions and simplifications of reality and take into account neither additional internal factors influencing language processing, such as attentional state or focus (Lappin and Leass, 1994; Lewis et al., 2006), emotion (Belavkin et al., 1999), and motivation (Belavkin, 2001), nor external factors such as visual context (Tanenhaus et al., 1995). Once a model has found support for underlying mechanisms of sentence processing, it can be used as a basis for investigating the effects of these additional factors. Therefore, the models discussed can be seen as a first step toward investigating such factors in the future.

A second limitation is related to a concern that Lewis and Vasishth (2005) raised: the degrees of freedom in cognitive models. For any set of cognitive models to be optimally comparable, they should be restricted by the same cognitive resources. However, cognitive architectures provide much freedom regarding different parameters (for example, the memory decay parameter in ACT-R can be changed manually). Therefore, models should generally strive to keep the quantitative parameters constant. If this is done, any variation between models will originate from the production rules and the content of the declarative memory, which is also where (linguistic) theory is implemented.

As a final limitation, any cognitive architecture that does not specify different types of memory (short-term memory, episodic long-term memory, semantic long-term memory) will make it difficult to model language processing in all its complexity. For example, long-term memory is difficult to implement in ACT-R,

because all chunks are subject to the same decay in activation over time. Thus, it is a puzzle why people do not forget certain pieces of knowledge that are not retrieved frequently (like, for example, what a hedgehog is). Recent research has found that different types of facts may actually have different decay rates (Sense et al., 2016). This can be important for language processing, because even infrequent words are not forgotten and can still be recognized and used after a long time. A related issue is that cognitive architectures with only one type of memory make it challenging to implement and manipulate working memory capacity. So, although the possibility of manipulating cognitive resources in cognitive models can be seen as a benefit, not restricting how these cognitive resources should be modeled limits its application. As language processing is known to be constrained by working memory capacity, manipulations of working memory capacity would be useful in order to study its effects on linguistic performance. Moreover, when modeling language acquisition or language attrition, working memory may be of great influence, as it can differ between ages (Grivol and Hage, 2011) and in clinical populations (e.g., ADHD: Martinussen et al., 2005; autism spectrum disorder: Barendse et al., 2013; cochlear implant users: AuBuchon et al., 2015). Although the function of working memory can be simulated indirectly through other processes like spreading activation (Daily et al., 2001), restrictions on their implementation in the cognitive architecture would make models more comparable and potentially more cognitively plausible.

Thus, using a cognitive architecture to investigate theories of linguistic competence has clear benefits as well as a number of current limitations. The main question in this review was to what extent general cognitive resources influence concretely implemented models of linguistic competence. An examination of the different cognitive models of linguistic performance provides evidence that well-studied general cognitive resources such as working memory influence language processing. In addition, less well-studied cognitive factors may also play a role, such as number of processing steps (Lewis and Vasishth, 2005) and processing efficiency (Van Rij et al., 2010). The influence of these factors can differ due to differences in, for example, experience, processing strategy, or possibly developmental disorder. Thus, our investigation of different cognitive models emphasizes that not only memory-related resources but also other timing-related resources and factors influence language processing.

As stated, implementations of linguistic theories into a cognitive model can, on the one hand, provide information about whether the theory can sufficiently account for observed performance. On the other hand, they can also be used to investigate

cognitive processes. For example, the speed of language processing is so high that it may not be met by the time-consuming processing steps provided by a cognitive model (cf. Vogelzang, 2017), or by the same memory processes that underlie other cognitive processes. So, from the viewpoint of linguistics, but also from the viewpoint of cognitive modeling, the puzzle of highly fast and efficient language processing compared to other cognitive processes is an interesting direction for future research.

Overall, cognitively constrained models can be used to investigate whether a linguistic theory can account for specific linguistic data. The interactions between a particular linguistic approach and general cognitive resources can be investigated through such models, which formalize of relation between competence and performance. Additionally, cognitive models can generate quantitative predictions of the basis of theories of linguistic competence. Because of this, cognitive models of linguistic theories are very suitable for investigating the relation between data, theory and experiments. Moreover, the possibility to model differences in cognitive resources allows for the investigation of individual differences in performance, as well as deviating performance due to aging or developmental disorders. In some cases, the high efficiency of language processing is currently not met by some of the constraining assumptions about cognitive resources. In this sense, cognitive models of language processing can also be used to investigate human cognition, for example in which ways currently adopted cognitive assumptions fail to meet the requirements for language processing. In conclusion, investigating specific linguistic phenomena through cognitive modeling can provide new insights that can complement findings from standard experimental techniques.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## ACKNOWLEDGMENTS

The authors would like to thank Michael Putnam for his comments on an earlier draft of the paper.

## FUNDING

This research was supported by a grant from the Netherlands Organization for Scientific Research (NWO) awarded to Jacolien Van Rij (grant no. 275-70-044). David Reitter acknowledges support from the National Science Foundation grant BCS-1734304.

## REFERENCES

- Aho, A. V., and Ullman, J. D. (1972). *The Theory of Parsing, Translation, and Compiling*. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Allen, R., and Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *J. Mem. Lang.* 55, 64–88. doi:10.1016/j.jml.2006.02.002
- Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York, NY: Oxford University Press.

- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004). An integrated theory of mind. *Psychol. Rev.* 111, 1036–1060. doi:10.1037/0033-295X.111.4.1036
- Arslan, B., Taatgen, N. A., and Verbrugge, R. (2017). Five-year-olds' systematic errors in second-order false belief tasks are due to first-order theory of mind strategy selection: a computational modeling study. *Front. Psychol.* 8:275. doi:10.3389/fpsyg.2017.00275
- AuBuchon, A. M., Pisoni, D. B., and Kronenberger, W. G. (2015). Short-term and working memory impairments in early-implanted, long-term cochlear implant

- users are independent of audibility and speech production. *Ear Hear.* 36, 733–737. doi:10.1097/AUD.0000000000000189
- Baauw, S. (2002). *Grammatical Features and the Acquisition of Reference. A Comparative Study of Dutch and Spanish*. New York: Routledge.
- Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997). Singulars and plurals in Dutch: evidence for a parallel dual-route model. *J. Mem. Lang.* 37, 94–117. doi:10.1006/jmla.1997.2509
- Barendse, E. M., Hendriks, M. P., Jansen, J. F., Backes, W. H., Hofman, P. A., Thoonen, G., et al. (2013). Working memory deficits in high-functioning adolescents with autism spectrum disorders: neuropsychological and neuroimaging correlates. *J. Neurodev. Disord.* 5, 14. doi:10.1186/1866-1955-5-14
- Belavkin, R. V. (2001). “Modelling the inverted-U effect in ACT-R,” in *Proceedings of the 2001 Fourth International Conference on Cognitive Modelling*, eds E. M. Altmann, A. Cleeremans, C. D. Schunn, and W. D. Gray (Mahwah, NJ, London: Lawrence Erlbaum), 275–276.
- Belavkin, R. V., Ritter, F. E., and Elliman, D. G. (1999). “Towards including simple emotions in a cognitive architecture in order to fit children’s behaviour better,” in *Proceedings of the 1999 Conference of the Cognitive Science Society (CogSci)* (Mahwah, NJ: Erlbaum).
- Belletti, A., and Guasti, M. T. (2015). *The Acquisition of Italian: Morphosyntax and Its Interfaces in Different Modes of Acquisition*. Amsterdam, PA: John Benjamins Publishing Company.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cogn. Psychol.* 18, 355–387. doi:10.1016/0010-0285(86)90004-6
- Bock, J. K., Dell, G. S., Chang, F., and Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition* 104, 437–458. doi:10.1016/j.cognition.2006.07.003
- Bock, J. K., and Griffin, Z. (2000). The persistence of structural priming: transient activation or implicit learning? *J. Exp. Psychol. Gen.* 129, 177–192. doi:10.1037/0096-3445.129.2.177
- Bock, J. K., and Kroch, A. S. (1989). “The isolability of syntactic processing,” in *Linguistic Structure in Language Processing*, eds G. N. Carlson and M. K. Tanenhaus (Dordrecht, The Netherlands: Springer), 157–196.
- Borst, J. P., Taatgen, N. A., and Van Rijn, H. (2010). The problem state: a cognitive bottleneck in multitasking. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 363–382. doi:10.1037/a0018106
- Branigan, H. P., Pickering, M. J., and Cleland, A. A. (1999). Syntactic priming in language production: evidence for rapid decay. *Psychon. Bull. Rev.* 6, 635–640. doi:10.3758/BF03212972
- Branigan, H. P., Pickering, M. J., Stewart, A. J., and McLean, J. F. (2000). Syntactic priming in spoken production: linguistic and temporal interference. *Mem. Cognit.* 28, 1297–1302. doi:10.3758/BF03211830
- Brasoveanu, A., and Dotlačil, J. (2015). Incremental and predictive interpretation: experimental evidence and possible accounts. *Proc. SALT* 25, 57–81. doi:10.3765/salt.v25i0.3047
- Bresnan, J. (2001). “The emergence of the unmarked pronoun,” in *Optimality-Theoretic Syntax*, eds G. Legendre, J. Grimshaw, and S. Vikner (Cambridge, MA: The MIT Press), 113–142.
- Budiu, R., and Anderson, J. R. (2002). Comprehending anaphoric metaphors. *Mem. Cognit.* 30, 158–165. doi:10.3758/BF03195275
- Bybee, J., and Beckner, C. (2009). “Usage-based theory,” in *The Oxford Handbook of Linguistic Analysis*, eds B. Heine and H. Narrog (Oxford: Oxford University Press), 827–855.
- Chang, F., Bock, J. K., and Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition* 99, 29–49. doi:10.1016/S0010-0277(03)00123-9
- Chang, F., Dell, G. S., and Bock, K. (2006). Becoming syntactic. *Psychol. Rev.* 113, 234–272. doi:10.1037/0033-295X.113.2.234
- Chesi, C. (2015). On directionality of phrase structure building. *J. Psycholinguist. Res.* 44, 65–89. doi:10.1007/s10936-014-9330-6
- Chien, Y.-C., and Wexler, K. (1990). Children’s knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Lang. Acquis.* 1, 225–295. doi:10.1207/s15327817la0103\_2
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1970). “Remarks on nominalization,” in *Reading in English Transformational Grammar*, eds R. Jacobs and P. Rosenbaum (Waltham, MA: Ginn), 184–221.
- Chomsky, N. (1980). Rules and representations. *Behav. Brain Sci.* 3, 1–15. doi:10.1017/S0140525X00001515
- Chomsky, N. (1981). *Lectures on Government and Binding: The Pisa Lectures*. Dordrecht, The Netherlands: Foris Publications.
- Chomsky, N. (1993). “A minimalist program for linguistic theory,” in *Tile View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, eds K. Hale and S. J. Keyser (Cambridge, MA: MIT Press), 1–52.
- Christiansen, M., and Chater, N. (2016). The now-or-never bottleneck: a fundamental constraint on language. *Brain Behav. Sci.* 39, e62. doi:10.1017/S0140525X1500031X
- Cleland, A. A., and Pickering, M. J. (2003). The use of lexical and syntactic information in language production: evidence from the priming of noun-phrase structure. *J. Mem. Lang.* 49, 214–230. doi:10.1016/S0749-596X(03)00060-3
- Corley, M., and Scheepers, C. (2002). Syntactic priming in English sentence production: categorical and latency evidence from an internet-based study. *Psychon. Bull. Rev.* 9, 126–131. doi:10.3758/BF03196267
- Daily, L. Z., Lovett, M. C., and Reder, L. M. (2001). Modeling individual differences in working memory performance: a source activation account. *Cogn. Sci.* 25, 315–353. doi:10.1016/S0364-0213(01)00039-8
- De Villiers, J., Cahillane, J., and Altruter, E. (2006). “What can production reveal about principle B?” in *Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition – North America*, Vol. 1, eds K. Deen, J. Nomura, B. Schulz, and B. Schwartz (Honolulu, HI: University of Connecticut), 89–100.
- Deese, J., and Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *J. Exp. Psychol.* 54, 180–187. doi:10.1037/h0040536
- Engelmann, F., Vasishth, S., Engbert, R., and Kliegl, R. (2013). A framework for modeling the interaction of syntactic processing and eye movement control. *Top. Cogn. Sci.* 5, 452–474. doi:10.1111/tops.12026
- Fedorenko, E., Woodbury, R., and Gibson, E. (2013). Direct evidence of memory retrieval as a source of difficulty in non-local dependencies in language. *Cogn. Sci.* 37, 378–394. doi:10.1111/cogs.12021
- Ferreira, V. S. (2003). The persistence of optional complementizer production: why saying “that” is not saying “that” at all. *J. Mem. Lang.* 48, 379–398. doi:10.1016/S0749-596X(02)00523-5
- Fillmore, C. J., Kay, P., and O’Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: the case of let alone. *Language* 64, 501–538. doi:10.2307/414531
- Fodor, J. A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Frazier, L., and Fodor, J. D. (1978). The sausage machine: a new two-stage parsing model. *Cognition* 6, 291–325. doi:10.1016/0010-0277(78)90002-1
- Fromkin, V. A. (2000). *Linguistics: An Introduction to Linguistic Theory*. Hoboken, NJ: Wiley-Blackwell Publishing.
- Gibson, E. (2000). “The dependency locality theory: a distance-based theory of linguistic complexity,” in *Image, Language, Brain*, eds A. Marantz, Y. Miyashita, and W. O’Neil (Cambridge, MA: MIT Press), 95–126.
- Grivol, M. A., and Hage, S. R. d. V. (2011). Phonological working memory: a comparative study between different age groups. *J. Soc. Bras. Fonoaudiol.* 23, 245–251. doi:10.1590/S2179-64912011000300010
- Grodner, D., and Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cogn. Sci.* 29, 261–290. doi:10.1207/s15516709cog0000\_7
- Grosz, B. J., Weinstein, S., and Joshi, A. K. (1995). Centering: a framework for modeling the local coherence of discourse. *Comput. Linguist.* 21, 203–225.
- Hale, J. T. (2011). What a rational parser would do. *Cogn. Sci.* 35, 399–443. doi:10.1111/j.1551-6709.2010.01145.x
- Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., and Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: evidence from written and spoken dialogue. *J. Mem. Lang.* 58, 214–238. doi:10.1016/j.jml.2007.07.003
- Hartsuiker, R. J., and Kolk, H. H. J. (1998). Syntactic persistence in Dutch. *Lang. Speech* 41, 143–184. doi:10.1177/002383099804100202
- Hartsuiker, R. J., and Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition* 75B, 27–39. doi:10.1016/S0010-0277(99)00080-3
- Hendriks, P. (2014). *Asymmetries between Language Production and Comprehension*. Dordrecht, The Netherlands: Springer.

- Hendriks, P., and Spender, J. (2006). When production precedes comprehension: an optimization approach to the acquisition of pronouns. *Lang. Acquis.* 13, 319–348. doi:10.1207/s15327817la1304\_3
- Hendriks, P., Van Rijn, H., and Valkenier, B. (2007). Learning to reason about speakers' alternatives in sentence comprehension: a computational account. *Lingua* 117, 1879–1896. doi:10.1016/j.lingua.2006.11.008
- Huetting, F., and Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Lang. Cogn. Neurosci.* 31, 80–93. doi:10.1080/23273798.2015.1047459
- Huttenlocher, J., Vasilyeva, M., and Shimpi, P. (2004). Syntactic priming in young children. *J. Mem. Lang.* 50, 182–195. doi:10.1016/j.jml.2003.09.003
- Jaeger, T. F., and Snider, N. E. (2008). "Implicit learning and syntactic persistence: surprisal and cumulativity," in *Proceedings of the 30th Annual Conference of the Cognitive Science Society (CogSci)* (Washington, DC: Cognitive Science Society), 1061–1066.
- Jaeger, T. F., and Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition* 127, 57–83. doi:10.1016/j.cognition.2012.10.013
- Jäger, L. A., Engelmann, F., and Vasisht, S. (2015). Retrieval interference in reflexive processing: experimental evidence from Mandarin, and computational modeling. *Front. Psychol.* 6:617. doi:10.3389/fpsyg.2015.00617
- Joshi, A. K., Levy, L. S., and Takahashi, M. (1975). Tree adjunct grammars. *J. Comput. Syst. Sci.* 10, 136–163. doi:10.1016/S0022-0000(75)80019-5
- Kaschak, M. P., Kutta, T. J., and Jones, J. L. (2011a). Structural priming as implicit learning: cumulative priming effects and individual differences. *Psychon. Bull. Rev.* 18, 1133–1139. doi:10.3758/s13423-011-0157-y
- Kaschak, M. P., Kutta, T. J., and Schatschneider, C. (2011b). Long-term cumulative structural priming persists for (at least) one week. *Mem. Cognit.* 39, 381–388. doi:10.3758/s13421-010-0042-3
- Kempen, G., and Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cogn. Sci.* 11, 201–258. doi:10.1207/s15516709cog1102\_5
- Kieras, D. E., and Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Hum. Comput. Interact.* 12, 391–438. doi:10.1207/s15327051hci1204\_4
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition* 2, 15–47. doi:10.1016/0010-0277(72)90028-5
- King, J., and Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *J. Mem. Lang.* 30, 580–602. doi:10.1016/0749-596X(91)90027-H
- Kong, F., Zhou, G., and Zhu, Q. (2009). "Employing the centering theory in pronoun resolution from the semantic perspective," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (Singapore: Association for Computational Linguistics), 987–996.
- Kootstra, G. J., Van Hell, J. G., and Dijkstra, T. (2012). Priming of code-switches in sentences: the role of lexical repetition, cognates, and language proficiency. *Bilingualism Lang. Cogn.* 15, 797–819. doi:10.1017/S136672891100068X
- Kuijper, S. J. M. (2016). *Communication Abilities of Children with ASD and ADHD*. Doctoral dissertation, University of Groningen, Groningen, The Netherlands.
- Kuijper, S. J. M., Hartman, C. A., and Hendriks, P. (2015). Who is he? Children with ASD and ADHD take the listener into account in their production of ambiguous pronouns. *PLoS ONE* 10:e0132408. doi:10.1371/journal.pone.0132408
- Lappin, S., and Leass, H. J. (1994). An algorithm for pronominal anaphora resolution. *Comput. Linguist.* 20, 535–561.
- Lewis, R. L. (1996). Interference in short-term memory: the magical number two (or three) in sentence processing. *J. Psycholinguist. Res.* 25, 93–115. doi:10.1007/BF01708421
- Lewis, R. L., and Vasisht, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.* 29, 375–419. doi:10.1207/s15516709cog0000\_25
- Lewis, R. L., Vasisht, S., and Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends Cogn. Sci.* 10, 447–454. doi:10.1016/j.tics.2006.08.007
- Malhotra, G. (2009). *Dynamics of Structural Priming*. Doctoral dissertation, University of Edinburgh, Edinburgh.
- Martinussen, R., Hayden, J., Hogg-Johnson, S., and Tannock, R. (2005). A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry* 44, 377–384. doi:10.1097/01.chi.0000153228.72591.73
- McKee, C. (1992). A comparison of pronouns and anaphors in Italian and English acquisition. *Lang. Acquis.* 2, 21–54. doi:10.1207/s15327817la0201\_2
- Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.
- Osborne, T., Putnam, M., and Gross, T. M. (2011). Bare phrase structure, label-less trees, and specifier-less syntax. Is minimalism becoming a dependency grammar? *Linguist. Rev.* 28, 315–364. doi:10.1515/tlir.2011.009
- Pauls, F., Macha, T., and Petermann, F. (2013). U-shaped development: an old but unsolved problem. *Front. Psychol.* 4:301. doi:10.3389/fpsyg.2013.00301
- Philip, W., and Coopmans, P. (1996). "The role of lexical feature acquisition in the development of pronominal anaphora," in *Amsterdam Series on Child Language Development*, Vol. 5, eds W. Philip and F. Wijnen (Amsterdam, The Netherlands: University of Amsterdam), 73–106.
- Pickering, M. J., and Branigan, H. P. (1998). The representation of verbs: evidence from syntactic priming in language production. *J. Mem. Lang.* 39, 633–651. doi:10.1006/jmla.1998.2592
- Pickering, M. J., and Ferreira, V. S. (2008). Structural priming: a critical review. *Psychol. Bull.* 134, 427–459. doi:10.1037/0033-2909.134.3.427
- Pollard, C., and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. Chicago, IL: University of Chicago Press.
- Prince, A., and Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical Report CU-CS-696-93, Department of Computer Science, University of Colorado at Boulder, and Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ. Published by Blackwell in 2004.
- Reitter, D. (2008). *Context Effects in Language Production: Models of Syntactic Priming in Dialogue Corpora*. Doctoral dissertation, University of Edinburgh, Edinburgh.
- Reitter, D., Keller, F., and Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cogn. Sci.* 35, 587–637. doi:10.1111/j.1551-6709.2010.01165.x
- Reitter, D., and Lebiere, C. (2010). "Accountable modeling in ACT-UP, a scalable, rapid-prototyping ACT-R implementation," in *Proceedings of the 10th International Conference on Cognitive Modeling (ICCM)* (Philadelphia, PA), 199–204.
- Reitter, D., and Moore, J. D. (2014). Alignment and task success in spoken dialogue. *J. Mem. Lang.* 76, 29–46. doi:10.1016/j.jml.2014.05.008
- Reitter, D., Moore, J. D., and Keller, F. (2006). "Priming of syntactic rules in task-oriented dialogue and spontaneous conversation," in *Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci)* (Vancouver, Canada: Cognitive Science Society), 685–690.
- Roberts, S., and Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychol. Rev.* 107, 358–367. doi:10.1037/0033-295X.107.2.358
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition* 42, 107–142. doi:10.1016/0010-0277(92)90041-F
- Salvucci, D. D., and Taatgen, N. A. (2008). Threaded cognition: an integrated theory of concurrent multitasking. *Psychol. Rev.* 115, 101–130. doi:10.1037/0033-295X.115.1.101
- Scheepers, C. (2003). Syntactic priming of relative clause attachments: persistence of structural configuration in sentence production. *Cognition* 89, 179–205. doi:10.1016/S0010-0277(03)00119-7
- Scheepers, C., Raffray, C. N., and Myachykov, A. (2017). The lexical boost effect is not diagnostic of lexically-specific syntactic representations. *J. Mem. Lang.* 95, 102–115. doi:10.1016/j.jml.2017.03.001
- Schoonbaert, S., Hartsuiker, R. J., and Pickering, M. J. (2007). The representation of lexical and syntactic information in bilinguals: evidence from syntactic priming. *J. Mem. Lang.* 56, 153–171. doi:10.1016/j.jml.2006.10.002
- Sense, F., Behrens, F., Meijer, R. R., and Van Rijn, H. (2016). An individual's rate of forgetting is stable over time but differs across materials. *Top. Cogn. Sci.* 8, 305–321. doi:10.1111/tops.12183
- Snider, N. E. (2008). *An Exemplar Model of Syntactic Priming*. Doctoral dissertation, Stanford University, Stanford.
- Spender, J., Smits, E.-J., and Hendriks, P. (2009). Coherent discourse solves the pronoun interpretation problem. *J. Child Lang.* 36, 23–52. doi:10.1017/S0305000908008854
- Steedman, M. (1988). "Combinators and grammars," in *Categorial Grammars and Natural Language Structures*, eds R. T. Oehrle, E. Bach, and D. Wheeler (Dordrecht, The Netherlands: Springer), 417–442.

- Steedman, M. (2000). *The Syntactic Process*. Cambridge, MA: MIT Press.
- Stewart, T., Tripp, B., and Eliasmith, C. (2009). Python scripting in the Nengo simulator. *Front. Neuroinformatics* 3:7. doi:10.3389/neuro.11.007.2009
- Stocco, A., and Crescentini, C. (2005). Syntactic comprehension in agrammatism: a computational model. *Brain Lang.* 95, 127–128. doi:10.1016/j.bandl.2005.07.069
- Taatgen, N., and Anderson, J. R. (2002). Why do children learn to say “broke”? A model of learning the past tense without feedback. *Cognition* 86, 123–155. doi:10.1016/S0010-0277(02)00176-2
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634. doi:10.1126/science.7777863
- Van Beijsterveldt, L. M., and Van Hell, J. G. (2009). Structural priming of adjective-noun structures in hearing and deaf children. *J. Exp. Child Psychol.* 104, 179–196. doi:10.1016/j.jecp.2009.05.002
- Van Maanen, L., and Van Rijn, H. (2010). The locus of the Gratton effect in picture-word interference. *Top. Cogn. Sci.* 2, 168–180. doi:10.1111/j.1756-8765.2009.01069.x
- Van Rijn, J. (2012). *Pronoun Processing: Computational, Behavioral, and Psychophysiological Studies in Children and Adults*. Doctoral dissertation, University of Groningen, Groningen, The Netherlands.
- Van Rijn, J., Van Rijn, H., and Hendriks, P. (2010). Cognitive architectures and language acquisition: a case study in pronoun comprehension. *J. Child Lang.* 37, 731–766. doi:10.1017/S0305000909990560
- Van Rijn, J., Van Rijn, H., and Hendriks, P. (2013). How WM load influences linguistic processing in adults: a computational model of pronoun interpretation in discourse. *Top. Cogn. Sci.* 5, 564–580. doi:10.1111/tops.12029
- Vogelzang, M. (2017). *Reference and Cognition: Experimental and Computational Cognitive Modeling Studies on Reference Processing in Dutch and Italian*. Doctoral dissertation, University of Groningen, Groningen, Netherlands.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Vogelzang, Mills, Reitter, Van Rijn, Hendriks and Van Rijn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Strong Generative Capacity and the Empirical Base of Linguistic Theory

Dennis Ott\*

Department of Linguistics, University of Ottawa, Ottawa, ON, Canada

## OPEN ACCESS

### Edited by:

Ángel J. Gallego,  
Universitat Autònoma de Barcelona,  
Spain

### Reviewed by:

Naoki Fukui,  
Sophia University, Japan  
Ricardo Etxepare,  
UMR5478 Centre de Recherche sur la  
Langue et les Textes Basques (IKER),  
France  
Norbert Hornstein,  
University of Maryland, College Park,  
United States

### \*Correspondence:

Dennis Ott  
dennis.ott@post.harvard.edu

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

Received: 29 April 2017

Accepted: 04 September 2017

Published: 21 September 2017

### Citation:

Ott D (2017) Strong Generative  
Capacity and the Empirical Base of  
Linguistic Theory.  
Front. Psychol. 8:1617.  
doi: 10.3389/fpsyg.2017.01617

This *Perspective* traces the evolution of certain central notions in the theory of Generative Grammar (GG). The founding documents of the field suggested a relation between the grammar, construed as recursively enumerating an infinite set of sentences, and the idealized native speaker that was essentially equivalent to the relation between a formal language (a set of well-formed formulas) and an automaton that recognizes strings as belonging to the language or not. But this early view was later abandoned, when the focus of the field shifted to the grammar's strong generative capacity as recursive generation of hierarchically structured objects as opposed to strings. The grammar is now no longer seen as specifying a set of well-formed expressions and in fact necessarily constructs expressions of any degree of intuitive "acceptability." The field of GG, however, has not sufficiently acknowledged the significance of this shift in perspective, as evidenced by the fact that (informal and experimentally-controlled) observations about string acceptability continue to be treated as *bona fide* data and generalizations for the theory of GG. The focus on strong generative capacity, it is argued, requires a new discussion of what constitutes valid empirical evidence for GG beyond observations pertaining to weak generation.

**Keywords:** generative grammar, grammaticality, acceptability, evidence, methodology

## INTRODUCTION

There exists a contradiction between the near-universal acceptance of acceptability judgments as a source of data for Generative Grammar (GG) on the one hand and the theory's express focus on *strong generative capacity* on the other. While linguists agree on this focus, they nevertheless tend to uncritically assume that judgments of the acceptability of strings constitute data for GG. But this assumption is baseless, and a renewed discussion of GG's empirical basis is in order.

## EARLY IDEALIZATIONS: THE SPEAKER AS AN AUTOMATON

Chomsky (1955, *LSLT*) defined as the "primary concern" of syntactic theory "to determine the grammatical sentences of any given language [...]" (57). Chomsky (1957, *SS*) elaborates:

"The fundamental aim in the linguistic analysis of a language *L* is to separate the grammatical sequences which are sentences of *L* from the ungrammatical sequences which are not sentences of *L* [...]. The grammar of *L* will thus be a device that generates all of the grammatical sequences of *L* and none of the ungrammatical ones." (*SS*, 13).

The set of sequences so determined "corresponds to the 'intuitive sense of grammaticality' of the native speaker" (*LSLT*, 95); hence, "the sequences generated by the grammar as grammatical



sentences must be acceptable, in some sense, to the native speaker [...]” (*LSLT*, 101). The adequacy of a grammar can be assessed by “[determining] whether or not the sequences that it generates are actually grammatical, i.e., acceptable to a native speaker” (SS, 13). Consequently, “the linguist’s task [is] that of producing [...] a grammar [that generates] all and only the sentences of a language [...]” (SS, 85).

On this *Early View* (EV), the idealized native speaker is the human equivalent of an automaton in the theory of formal languages, which accepts (recognizes) or rejects a given string depending on whether or not it is part of the set of legal sequences. While the importance of hierarchical structures underlying the sequences was recognized to be of central importance, the formal systems used at the time—Post-style rewrite rules plus transformational rules—ultimately enumerated strings (see Lasnik, 2000).

*LSLT* and *SS* took grammaticality (or degrees thereof, as argued in *LSLT*: chapter 5; also Chomsky, 1965, p. 148ff.) to be accessible to intuition. That the matter is more complex was explicitly acknowledged shortly after by Chomsky (1965, p. 11), who cautions that “[t]he notion ‘acceptable’ is not to be confused with ‘grammatical’”: while the former “belongs to the study of performance,” the latter “belongs to the study of competence [...]” This is the standard distinction between grammaticality and acceptability, often not drawn properly even in technical papers (cf. Newmeyer, 1983, p. 51). A true shift in perspective, however, took place later, when the notion of *sentence*, understood as *sequence in L*, was eliminated altogether from the theory.

## A SHIFT IN PERSPECTIVE

Later works of Chomsky’s are explicit in rejecting the EV and its view of the idealized native speaker as a human automaton. Perhaps the first clear articulation of this shift appears in Chomsky (1980), where we find the assertion that “[a GG] does not in and of itself determine the class of what we might choose to call ‘grammatical sentences’ [...]” an unremarkable conclusion “once we recognize that the fundamental concepts are *grammar* and *knowing a grammar*, and that *language* and *knowing a language* are derivative” (p. 126).

This dismissal of the view of a *language* as a set of sentences is a corollary of the shift of the focus of attention from sentences to *structures*:

“For each sentence, the grammar determines aspects of its phonetic form, its meaning and perhaps more. [...] [It] is said to ‘weakly generate’ the sentences of the language and to ‘strongly generate’ the structural descriptions of these sentences” (Chomsky, 1980, p. 220).

The grammar *strongly* generates structural descriptions (SDs), not strings; the latter can at best be said to be generated in some *weak* sense, in that the “phonetic form” associated by the grammar with any SD has sequential properties (Chomsky, 1990). Importantly, the grammar is now no

longer taken to generate objects of which the property “acceptability” or “well-formedness” could be predicated (i.e., strings/sequences/sentences).

Chomsky goes further in suggesting that the focus on strong generative capacity (SGC) in fact *requires* the generation of “deviant” expressions, as a matter of empirical fact:

“[A] GG will not generate the set of sentences that a speaker-hearer will regard as acceptable; indeed, it is virtually a criterion of adequacy that it should not, since so many different factors enter into such judgments” (Chomsky, 1980, p. 274 fn. 54).

Chomsky (1986, p. 24) adds that “[a] GG is not a set of statements about externalized objects constructed in some manner;” to which he refers as “E(external)-language,” as opposed to the I(nternal)-language that constructs SDs underlying these objects (see already Chomsky, 1959, 1963). This move replaces the EV of the grammar as determining a set of *sentences* with one of grammar as determining *form-meaning correlations*:

“[When a person knows a language], we do not mean that he or she knows an infinite set of sentences [...]; rather, what we mean is that the person knows what makes sound and meaning relate to one another in a specific way [...]” (Chomsky, 1986, p. 27).

Consequently, it is “meaningless to ask whether [some intuitively “deviant” expression] is, or is not, a member of the E-language weakly generated by *L*; and nothing would follow from a discovery (or stipulation) one way or another” (Chomsky, 1990, p. 145).

Chomsky (1986, p. 29f.) explains the motivation for the EV with the influence of formal-language theory on the then-nascent field of GG, an analogy now explicitly dismissed:

“In the literature of [GG], the term ‘language’ has regularly been used for E-language in the sense of a set of well-formed sentences [...]. The misleading choice of terms was, in part [due to] the confluence of two intellectual traditions: traditional and structuralist grammar, and the study of formal systems. [...] But the study of formal languages was misleading in this regard. When we study [a formal language], we may take it to be a ‘given’ [...] infinite class of sentences in some given notation. Certain expressions in this notation are well-formed sentences, others are not. [...] It is easy to see how one might take over from the study of formal languages the idea that the ‘language’ is somehow given as a set of sentences [...], while the grammar is some characterization of this infinite set [...]. The move is understandable, but misguided; [...] the E-language is not ‘given.’”

Chomsky and Lasnik (1993, p. 508) reiterate this dismissal of the EV:

“[A] ‘formal language’ in the technical sense [is] a set of well-formed formulas [...]. Call such a set an E-language [...]. In the theory of formal languages, the E-language is defined by

stipulation, hence is unproblematic. But it is a question of empirical fact whether [...] I-language generates not only a set of [structures] but also a distinguished E-language [...]. [T]he concept of E-language [...] has no known status in the study of language [...].”

The field of GG ostensibly followed Chomsky in shifting the focus from strings to SDs and their properties. What is customarily ignored, however, is that such a shift leaves notions such as “acceptability” or “well-formed sentence” with no immediate relevance to the theory of SGC.

## BEYOND ACCEPTABILITY

Chomsky (1986, p. 98ff.) illustrates the practical effects of this shift in focus with a concrete example. While (1), where *who* is displaced from the gap position, receives a straightforward interpretation in terms of an operator-variable dependency, (2) cannot be interpreted in this way.

- (1) (I know) [who [John [kissed \_]]]  
 (“which person *x* is such that John kissed *x*”)  
 (2) (I know) [who [John [kissed Mary]]]

In (2), the *wh*-operator has no variable to bind, and consequently cannot be assigned an interpretation. Importantly, we cannot simply “neglect” the fronted *wh*-phrase and interpret (2) as meaning (*I know*) *John kissed Mary*, a fact that Chomsky attributes to the principle of Full Interpretation—an interface condition, in current parlance. Does this mean that we want to block generation of (2), while allowing generation of (1)? Chomsky explicitly denies this, arguing that such a move would redundantly replicate the effect of Full Interpretation. Consequently, both SDs in (1) and (2) are grammatical (generated by the grammar); the “deviance” of (2) is due to an extraneous principle of interpretation. But the fact that the string deriving from (2) is “deviant” *per se* is of no immediate concern to the theory of grammar.

Analogously, to use the famous example introduced in *LSLT* (145), the goal of the theory is not to construct a grammar that generates a set of well-formed formulas including *Colorless green ideas sleep furiously* but excluding *Furiously sleep ideas green colorless*, but to explain why the SD assigned to the latter cannot be mapped onto an analogous interpretation. The naturalness of the typographical or acoustic object is of no immediate relevance to the theorist (cf. McCawley, 1982, p. 78f.). Similarly, island constraints are not generalizations over classes of sentences that are “unacceptable,” but describe the absence of otherwise expectable interpretations of expressions. The fact that *What does John like apples and?* is an intuitively “unacceptable” string is a mere observation; what does constitute a relevant explanandum for GG is solely the fact that it unexpectedly fails to mean “which *x* is such that John like apples and *x*?” (*pace* Preminger, in press).

On this *Revised View* (RV), the empirical success of GG depends on its ability to correctly model the speaker’s knowledge of sound-meaning relations, not the intuitive acceptability of strings:

“Linguistic expressions may be ‘deviant’ along all sorts of incommensurable dimensions, and we have no notion of ‘well-formed sentence’ [...]. Expressions have the interpretations assigned to them by the performance systems in which the language is embedded: period” (Chomsky, 1993, p. 27).

In later works, Chomsky entertains the idea that generation of SDs proceeds freely via the operation Merge, with constraints imposed only by external systems. For instance, Chomsky (2004, p. 111) argues that “theta-theoretic failures at the interface do not cause the derivation to crash; such structures yield ‘deviant’ interpretations of a great many kinds.” The relevant “theta-theoretic failures” are interface properties of SDs that are strongly generated, regardless of the deviance of derivative stimuli they may incur. More generally:

“Merge can apply freely, yielding expressions interpreted at the interface in many different kinds of ways. They are sometimes called ‘deviant,’ but that is only an informal notion. [...] The only empirical requirement is that [the interfacing systems] assign the interpretations that the expression actually has, including many varieties of ‘deviance’” (Chomsky, 2008, p. 144).

Chomsky (2016, p. 3f.) notes that “[f]ree application of rules can yield deviant expressions, but that is unproblematic, in fact required. Deviant expressions should be generated with their interpretations [...],” as “[i]t would radically complicate the generative procedure if [Merge] were required to yield non-deviant structures,” “even assuming that the concept [of deviance, D.O.] can be defined in absolute terms, which has never been obvious” (fn. 8).

On this RV, there exists no notion of well-formedness that is given independently of whatever is strongly generated by the I-language. The grammar does not specify a set of legal strings but an infinity of SDs; the only empirical success criterion is that the SDs postulated by the theorist have the properties in interpretation and externalization they do.

## QUO VADIS?

While the field ostensibly embraced the focus on SGC and SDs championed by Chomsky, the EV remains widely adopted in actual practice. Grammaticality and acceptability are standardly equated, and I-languages taken to determine sets of well-formed strings/sentences. The following quotes, randomly culled from popular textbooks, are representative:

“We say that an utterance is *grammatical* if native speakers judge it to be a possible sentence of their language” (O’Grady and Archibald, 2016, p. 139).

“The psychological experiment used to get to [the speaker’s knowledge of language] is called the grammaticality-judgment task. The judgment task involves asking a native speaker to read a sentence, and judge whether it is well-formed (grammatical), marginally well-formed, or ill-formed (unacceptable or ungrammatical)” (Carnie, 2013, p. 14).

“[A] sequence of words is called a string. Putting a star at the start of a string is a claim that it isn’t a grammatical sentence of the language in question” (Adger, 2003, p. 4).

“A [...] reason for using grammaticality judgments [*sic*] is to obtain a form of information that scarcely exists within normal language use at all—namely, negative information, in the form of strings that are not part of the language” (Schütze, 1996, p. 2).

In a survey of empirical methods, Schütze (2011, p. 207) identifies the assumption “that our mental grammar distinguishes at least two kinds of strings: those that are possible sentences of our language and those that are not” as “Chomsky’s view,” despite the fact that Chomsky has defended the opposite for at least 40 years.

As a result of this (unconscious?) adherence to the EV, acceptability judgments continue to take center stage in GG, and a good deal of the literature on experimental syntax has been devoted to refining their elicitation (Sprouse, 2013). Sprouse (2007, p. 123) notes that experimental methods have made it “almost trivial to detect subtle differences along a continuous spectrum of acceptability,” which he takes to raise the question of “whether the working assumption of the past 40 years should be abandoned”—this being the assumption “that grammatical knowledge is categorical—sentences are either grammatical or ungrammatical.” He explains that “the psychological claim underlying theories of categorical grammaticality is that ungrammatical sentences have no licit representation, [i.e.] cannot be constructed from the available mental representations.” There is no recognition of the fact that there exists no notion of “(un)grammatical sentence” on the RV, or any argument to the contrary.

The above remarks illustrate that the profound implications of the RV and its focus on generation of SDs remain insufficiently appreciated (cf. Fukui, 2015), and that the field’s continuing obsession with string acceptability betrays the lasting impact of the EV. Technical work in GG remains strongly dominated by the assumption that syntactic computation ought to be virtually or entirely “crash-proof,” generating all and only those expressions that give rise to strings that are acceptable to the native speaker (*modulo* performance-related factors). This view is most explicitly espoused by Frampton and Gutmann (2002, p. 90), who maintain that “an optimal derivational system [...] is a system that generates only objects that are well-formed and satisfy conditions imposed by the interface systems.” Note the use of the term “objects,” intended to ambiguously cover both sentences (the focus of the EV) and SDs and their semantic and phonological correlates (the focus of the RV).

This conceptually confused fixation on “crash-proofness” has given rise to a plethora of proposals that enrich the syntactic machinery in order to avoid “overgeneration” (e.g., by blocking certain extractions), ignoring the fact that this notion has no obvious relevance on the RV. A direct outgrowth of this ideology is the extensive reliance on highly stipulative features as licensors of structure-building (Chomsky, 2001, p. 6), leading to a “highly baroque syntax” (Reinhart, 2006, p. 5) employing “diacritic features that have no detectable properties other than their ability to trigger [syntactic operations]” (Richards, 2016, p. 1). Space precludes further discussion of the technical literature here; see Ott and Šimík (in progress).

The methodological problem posed by acceptability judgments, no matter how experimentally refined, is not their informal and inherently behavioral nature (Bever, 1970), but the fact that they do not constitute explananda for a theory of I-language (as opposed to E-language). The shift from the EV to the RV, traced above, demands a focus on speakers’ knowledge of form-meaning correlations rather than string acceptability. Of course, in many cases “acceptability judgments” are in fact shorthand for judgments about such correlations—we can say that *He<sub>i</sub> likes John<sub>i</sub>* is “unacceptable,” or that it lacks the intended reading; we can say that (2) above is “deviant,” with an implicit understanding that we’re referring to the absence of an interpretation analogous to (1). This innocent informal usage aside, however, the “(un)acceptable” status of sentences remains the de-facto empirical benchmark for theoretical proposals within GG, and informal observations about weak generative capacity, clad in technical terms, are standardly elevated to the status of generalizations to be accounted for (cf. the case of islands mentioned above). The field must overcome these limitations and move on to a theoretical characterization of possible SDs (e.g., in terms of the theory of Merge) and their interface properties (Chomsky et al., 2017). This will require the recognition that fears of “overgeneration” are unfounded, and more generally that GG’s object of inquiry is much more abstract than the EV and its convenient idealizations suggested.

## CONCLUSION

The theory of GG has undergone significant conceptual shifts. Early work construed a GG as a finitary procedure that recursively enumerates all and only well-formed sentences of a language. Later work abandoned this conception entirely in favor of generation of discrete, hierarchically structured objects (I-language). Despite this shift, the field has retained a methodological obsession with the intuitive well-formedness of strings and associated notions such as “overgeneration” (E-language).

Chomsky (1965, p. 63) noted that “discussion of weak generative capacity marks only a very early and primitive stage of the study of [GG]. Questions of real linguistic interest arise only when [SGC] [...] becomes

the focus of discussion.” It is high time that this remark be taken seriously, which will necessitate a renewed discussion of the field’s goals and the question of which observations can be translated into valid explananda for the theory, as opposed to mere translation of these observations into technical vocabulary. This will likely require the incorporation of various forms of evidence, from introspective to neurological, that can be hoped to tap the human “notion of structure,” in Jespersen’s famous formulation.

## REFERENCES

- Adger, D. (2003). *Core Syntax*. Oxford: Oxford University Press.
- Bever, T. G. (1970). “The cognitive basis for linguistic structures,” in *Cognition and the Development of Language*, ed J. R. Hayes (New York, NY: John Wiley & Sons), 279–362.
- Carnie, A. (2013). *Syntax*. Malden, MA: Wiley-Blackwell.
- Chomsky, N. (1955). *The Logical Structure of Linguistic Theory*. Ms., Harvard University and MIT (Page references are to the revised 1975 version published by Plenum, New York, NY).
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton & Co.
- Chomsky, N. (1959). On certain formal properties of grammar. *Inform. Control* 2, 137–167.
- Chomsky, N. (1963). “Formal properties of grammars,” in *Handbook of Mathematical Psychology*, Vol. II, eds R. D. Luce, R. R. Bush, and E. Galanter (New York, NY: Wiley), 323–418.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1980). *Rules and Representations*. New York, NY: Columbia University Press.
- Chomsky, N. (1986). *Knowledge of Language*. New York, NY: Praeger.
- Chomsky, N. (1990). On formalization and formal linguistics. *Nat. Lang. Linguist. Theory* 8, 143–147.
- Chomsky, N. (1993). “A minimalist program for linguistic theory,” in *The View From Building 20*, eds K. Hale and S. J. Keyser (Cambridge, MA: MIT Press), 1–52.
- Chomsky, N. (2001). “Derivation by phase,” in *Ken Hale: A Life in Language*, ed M. Kenstowicz (Cambridge, MA: MIT Press), 1–52.
- Chomsky, N. (2004). “Beyond explanatory adequacy,” in *Structures and Beyond*, ed A. Belletti (Oxford: Oxford University Press), 104–131.
- Chomsky, N. (2008). “On phases,” in *Foundational Issues in Linguistic Theory*, eds R. Freidin, C. P. Otero, and M. L. Zubizarreta (Cambridge, MA: MIT Press), 291–321.
- Chomsky, N. (2016). “Puzzles about phases,” in *Linguistic Variation*, eds L. Franco and G. Bellucci (Berlin; New York, NY: Mouton de Gruyter).
- Chomsky, N., Gallego, Á., and Ott, D. (2017). *Generative Grammar and the Faculty of Language: Insights, Questions, and Challenges*. Ms., MIT, Universitat Autònoma de Barcelona, and University of Ottawa.
- Chomsky, N., and Lasnik, H. (1993). “The theory of principles and parameters,” in *Syntax*, eds J. Jacobs, A. von Stechow, W. Sternefeld, and T. Vennemann (Berlin; New York, NY: Walter de Gruyter), 506–569.
- Frampton, J., and Gutmann, S. (2002). “Crash-proof syntax,” in *Derivation and explanation in the Minimalist Program*, eds S. D. Epstein and T. D. Seely (Oxford: Blackwell), 90–105.
- Fukui, N. (2015). “A note on weak vs. strong generation in human language,” in *50 Years Later: Reflections on Chomsky’s Aspects*, eds Á. Gallego and D. Ott (Cambridge, MA: MITWPL), 125–131.
- Lasnik, H. (2000). *Syntactic Structures Revisited*. Cambridge, MA: MIT Press.
- McCawley, J. D. (1982). How far can you trust a linguist? in *Language, Mind, and Brain*, eds T. W. Simon and R. J. Scholes (Hillsdale, NJ: Erlbaum), 75–88.
- Newmeyer, F. (1983). *Grammatical Theory*. Chicago: University of Chicago Press.
- O’Grady, W., and Archibald J. (2016). *Contemporary Linguistic Analysis*. Toronto, ON: Pearson.
- Preminger, O. (in press). “Back to the Future: Non-generation, filtration, and the heartbreak of interface-driven minimalism,” in *Syntactic Structures after 60 Years: The Impact of the Chomskyan Revolution in Linguistics*, eds N. Hornstein, H. Lasnik, P. Patel-Grosz and C. Yang (Berlin: Mouton de Gruyter).
- Reinhart, T. (2006). *Interface Strategies*. Cambridge, MA: MIT Press.
- Richards, N. (2016). *Contiguity Theory*. Cambridge, MA: MIT Press.
- Schütze, C. T. (1996). *The Empirical Base of Linguistic Theory*. Chicago: University of Chicago Press (Page references are to the 2016 reprint by Berlin: Language Science Press).
- Schütze, C. T. (2011). Linguistic evidence and grammatical theory. *WIREs Cognitive Science* 2, 206–221. doi: 10.1002/wcs.102
- Sprouse, J. (2007). Continuous acceptability, categorial grammaticality, and experimental syntax. *Biolinguistics* 1, 123–134.
- Sprouse, J. (2013). “Acceptability judgments,” *Oxford Bibliographies*. Available online at: <http://www.oxfordbibliographies.com>

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## FUNDING

Publication of this article was partially funded by the University of Ottawa, whose financial support is gratefully acknowledged.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Ott. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Locus Preservation Hypothesis: Shared Linguistic Profiles across Developmental Disorders and the Resilient Part of the Human Language Faculty

Evelina Leivada<sup>1,2\*</sup>, Maria Kambanaros<sup>2,3</sup> and Kleanthes K. Grohmann<sup>2,4</sup>

<sup>1</sup> Language and Culture, UiT-The Arctic University of Norway, Tromsø, Norway, <sup>2</sup> Rehabilitation Sciences, Cyprus University of Technology, Limassol, Cyprus, <sup>3</sup> Cyprus University of Technology, Limassol, Cyprus, <sup>4</sup> English Studies, University of Cyprus, Nicosia, Cyprus

## OPEN ACCESS

### Edited by:

Ángel J. Gallego,  
Universitat Autònoma de Barcelona,  
Spain

### Reviewed by:

LouAnn Gerken,  
University of Arizona, United States  
F. Sayako Earle,  
University of Delaware, United States

### \*Correspondence:

Evelina Leivada  
evelina@biolinguistics.eu

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 15 July 2017

**Accepted:** 25 September 2017

**Published:** 13 October 2017

### Citation:

Leivada E, Kambanaros M and Grohmann KK (2017) The Locus Preservation Hypothesis: Shared Linguistic Profiles across Developmental Disorders and the Resilient Part of the Human Language Faculty. *Front. Psychol.* 8:1765. doi: 10.3389/fpsyg.2017.01765

Grammatical markers are not uniformly impaired across speakers of different languages, even when speakers share a diagnosis and the marker in question is grammaticalized in a similar way in these languages. The aim of this work is to demarcate, from a cross-linguistic perspective, the linguistic phenotype of three genetically heterogeneous developmental disorders: specific language impairment, Down syndrome, and autism spectrum disorder. After a systematic review of linguistic profiles targeting mainly English-, Greek-, Catalan-, and Spanish-speaking populations with developmental disorders ( $n = 880$ ), shared loci of impairment are identified and certain domains of grammar are shown to be more vulnerable than others. The distribution of impaired loci is captured by the *Locus Preservation Hypothesis* which suggests that specific parts of the language faculty are immune to impairment across developmental disorders. Through the Locus Preservation Hypothesis, a classical chicken and egg question can be addressed: Do poor conceptual resources and memory limitations result in an atypical grammar or does a grammatical breakdown lead to conceptual and memory limitations? Overall, certain morphological markers reveal themselves as highly susceptible to impairment, while syntactic operations are preserved, granting support to the first scenario. The origin of resilient syntax is explained from a phylogenetic perspective in connection to the “syntax-before-phonology” hypothesis.

**Keywords:** distributed morphology, grammatical marker, linguistic phenotype, syntax, Autism spectrum disorders (ASD), Down Syndrome, specific language impairment (SLI)

## INTRODUCTION

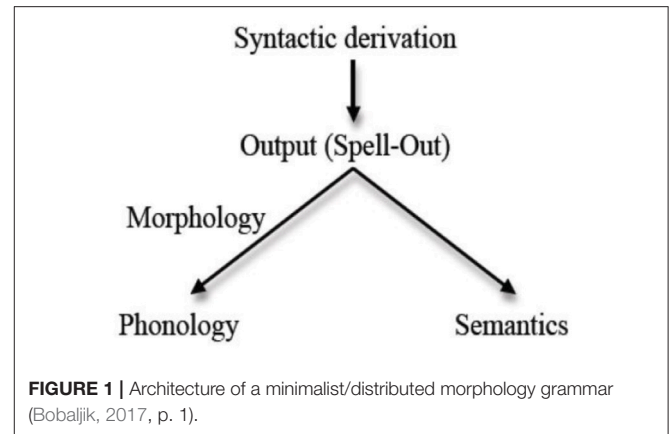
In his seminal book *The Biological Foundations of Language*, Eric Lenneberg made the following observation when comparing different states of verbal behavior:

Some aphasic symptoms bear certain similarities to the common derangements of speech and language seen in individuals in good health under conditions of mental exhaustion or states of drowsiness [...]. Clinically, we may encounter an almost kaleidoscopic combination of idiosyncratic failure or sparing of particular skills which renders precise correlations between pathological anatomy and pathological verbal behavior very difficult (Lenneberg, 1967, p. 222; emphasis added).

Almost 40 years later, Phillips (2005) observed, when comparing the underpinnings of various developmental language impairments, that some aspects of language such as morphosyntactic difficulties associated with tense inflection appear to be affected across pathologies with different genetic causes (e.g., Specific Language Impairment, autism, Williams Syndrome, Down Syndrome, fragile X syndrome). Similarly, many studies have identified overlaps at the phenotypic level among different disorders: Leyfer et al. (2008) and Durrleman and Delage (2016) for autism spectrum disorder (ASD) and specific language impairment (SLI), Perovic et al. (2013) for ASD and Williams Syndrome, Dykens et al. (2011) for Prader-Willi syndrome and ASD, Eadie et al. (2002) for Down Syndrome (DS) and SLI, and Bishop (2014b) for anterior aphasia and SLI.

Overlaps in the behavioral profile of populations with different diagnoses have led to the claim that variation across phenotypes (i.e., breakdowns) is constrained in a way that renders some aspects of language processing—or more generally, cognition—more vulnerable in all pathological conditions, while others are consistently spared across individuals and conditions, both acquired and developmental (Phillips, 2005; Glisky, 2007; Kambanaros and van Steenbrugge, 2013; Benítez-Burraco and Boeckx, 2014; Leivada, 2014, 2015; Kambanaros and Grohmann, 2015; Tsimpli et al., 2017a). This high vulnerability of certain aspects of language is possibly the result of brain network organization. Studies on the distribution of lesions at the human connectome suggest that hubs are more likely to be anatomically abnormal than non-hubs across many, or possibly all, brain disorders because of their high centrality (van den Heuvel and Sporns, 2013; Crossley et al., 2014). Furthermore, multiple theoretical perspectives and neuroimaging research are addressing outstanding questions about the nature and extent of brain connectivity aberrations in SLI vs. autism (Verhoeven et al., 2012) and DS (Anderson et al., 2013).

At the same time, studies disagree about the status of a grammatical marker as vulnerable or not, even when reporting on the competence and/or performance of speakers of the *same* language; for example, see Manika et al. (2010) for the greater variability that exists between studies that report on the status of clitics in Greek SLI. This phenotypic variability across linguistic profiles is observed even *within* one pedigree, where affected members share a diagnosis, as Bartha-Doering et al. (2016) have shown for SLI. One could, of course, argue that this is due to the character of SLI as a disorder that relies on an exclusionary diagnosis. In other words, because the criteria for diagnosing SLI are exclusionary (Reilly et al., 2014), this inevitably forms a largely heterogeneous disorder with diverse subtypes that encompass very different populations. However, the same phenotypic variability can be observed in impaired phenotypes that rely on an inclusionary diagnosis. For instance, Fowler (1995) notes that there is tremendous variability with respect to language function in individuals with DS. Lecavalier (2014) raises the same observation for ASD. Overall, this variability could be the result of variable expressivity. Individuals that carry a pathogenic variant of a gene can be impaired in a non-uniform fashion and this may result in different cognitive subtypes within an impaired phenotype (see



Geschwind, 2011 for a review of variable expressivity in ASD). In this context, it becomes clear that the attained performance is not necessarily homogeneous even among people that share the same developmental disorder and speak the same language.

The picture painted by this brief overview involves a paradox. Although specific markers are highly vulnerable and as such prone to impairment across disorders, there still exists a lot of variability in terms of the attested impairment both across and within disorders. Phillips (2005) calls this state of affairs “a clear puzzle” and presents it in the following way:

On the one hand, the effects of specific genetic disorders on language appear to be surprisingly *nonspecific*. Similar aspects of language appear to be impacted across a variety of disorders with different genetic causes. On the other hand, the effects of genetic disorders on language are *highly specific*. [Developmental language impairments] appear to *selectively target* certain subparts of language while sparing others (Phillips, 2005, p. 79; emphases added).

This picture might even include derangements of speech in healthy, neurotypical adults, as noted in Lenneberg (1967) and quoted above.

In the present work, it is argued that the solution to Phillips’ puzzle requires (i) a fine-grained analysis of loci of variation across different developmental impairments, which is (ii) situated within linguistic frameworks that put forth a clear division of labor between the different parts of grammar, and (iii) approached from a cross-linguistic perspective. In what follows we present work on (i), on (ii), and in parts on (iii), through comparing the linguistic profiles of three different types of developmental disorders (SLI, ASD, and DS) in speakers of two varieties of Greek (Standard Modern Greek and Cypriot Greek), English, Spanish, and Catalan.

We employ the layout of grammar put forth in the framework of Distributed Morphology (Halle and Marantz, 1993; Harley and Noyer, 1999; **Figure 1**) in order to identify which aspects of language feature the various loci of impairment. This model does not enhance the testability of our argument—but it does facilitate organizing the distribution of impaired markers across levels of linguistic analysis in a transparent way. In this framework (and minimalism at large), “syntactic derivation” refers to operations

in syntax proper, the outcome of which feeds the other levels of analysis: phonology (via Phonetic/Phonological Form) and semantics (via Logical Form). Spell-out is an instruction to transfer this outcome to the next stage of operations.

By using a theoretical linguistics model as a vehicle for cartographing vulnerable loci across disorders, we establish an interdisciplinary connection between theoretical linguistics and the clinical aspects of cognitive neuroscience. Such interdisciplinary bridges are crucial in the study of language perhaps today more than ever, for it has been recently argued that linguistics, once seen the key player in the field of cognitive science, has seen its influence on closely allied disciplines fade away over the last years (Ferreira, 2005; Hagoort, 2014). However, one should not ignore the considerable body of literature that establishes interdisciplinary bridges in a way that shows how notions and primitives from theoretical linguistics can contribute to the study of neuroscience and other closely allied disciplines (Marantz, 2005; Sprouse and Almeida, 2013; Leivada, 2015). Against this background, the second aim of the present work is to offer a concrete example of how models of grammar in theoretical linguistics can inform the study of the brain through the investigation of pathological phenotypes. The study of the latter offers a unique perspective into the “physical mechanisms of the brain that correspond to the various domains of grammar and its structure” (Terzi, 2005, p. 111).

## METHODS

The case reports presented in the following are the result of extensive database searching through PubMed, SCOPUS, ScienceDirect, and Google Scholar, as well as probing individual journals for results retrieved by searches for any combination of the terms “primary/specific language impairment,” “autism spectrum disorder(s),” “Down(s) syndrome,” “linguistic phenotype,” “impaired/atypical phonology/morphology/syntax/semantics/pragmatics,” “word retrieval in SLI/ASD/DS,” and “linguistic impairment/disorder.” Our searches were constrained in terms of a time frame that covered the last two decades and in terms of language groups (Greek, English, Catalan, and Spanish). In choosing these language combinations, our aim was to cover both monolingual (Standard Modern Greek, English, Spanish) and bilingual/bicultural populations (Standard Modern Greek–Cypriot Greek, Spanish–Catalan) and languages with rich morphology. A cross-linguistic perspective is likely to shed light to the vulnerable parts of language in a way that goes beyond language-specific particularities. If any, it is the cross-linguistic study of the pathologies under investigation that has the potential to uncover the common denominator and the factors that distinguish children with a pathological linguistic profile from their typically developing peers (Leonard, 2014).

### Specific Language Impairment (SLI)

Specific language impairment is a developmental disorder marked by limitations in the process of language development. It is usually assumed that these limitations occur in the absence of neurological damage such as hearing impairment, motor skills

disorder, and low non-verbal IQ, and in the presence of otherwise typical cognitive development (Leonard, 1998). SLI is largely heterogeneous and many distinct subtypes have been identified in the literature. Two common SLI subtypes are typical SLI and pragmatic language impairment (Bishop, 2004): the former refers to those cases that involve problems with grammatical development (e.g., omission of past tense morphemes in English), sometimes referred to as G(rammatical)-SLI (van der Lely, 2005) or Sy(ntactic)SLI (Friedmann and Novogrodsky, 2008), while the latter indicates social communication problems (e.g., lack of coherence in conversation). In some studies, these linguistic limitations have been grounded in cognition rather than language *per se* (e.g., working memory limitations; Gathercole and Baddeley, 1990; Dodwell and Bavin, 2008), leading to the conclusion that SLI is not really specific to language, as its name suggests (Engel de Abreu et al., 2014). This has led to serious debates, and no consensus, in the literature on terminology for defining SLI (Bishop, 2014a; Reilly et al., 2014). **Table 1** presents 44 studies that feature different language groups and sets of tasks. These studies have been selected so as to include representation of all domains of impairment that have been proposed in the relevant literature on SLI.

**Table 1** identifies all domains of grammar as potentially impaired in SLI populations. However, a closer look at the relevant results suggests that only *some* domains of grammar are truly atypical. It is clear that many studies report problems in morphophonology or pragmatics as well as general processing limitations. The nature of the impairment is less clear, though, in studies that argue in favor of a problem in the syntactic domain. Before showing why, we understand syntax as (the iterative application of) the operations (internal and external) Merge and Agree, following the definitions of Chomsky (2001). Many of the studies reviewed refer to omissions of agreement markers or failure to establish agreement/binding relations between different components of structure when talking about impaired syntax (e.g., Clahsen and Dalalakis, 1999; Tsimpli and Stavrakaki, 1999; Lin, 2007), and we follow the assumption that syntax indeed hosts these relations. The reason is that it is necessary to revisit the *results* of these studies—and explain in what sense they are not truly making a case for a deficient syntax—, instead of evoking an argument that dismisses the syntactic nature of these relations (binding/agreement) on theoretical grounds (e.g., by suggesting that Agree takes place post-syntactically, so when a study reports agreement errors, this does not concern syntax in the first place).

Returning to the studies in **Table 1**, Loeb et al. (1998) claim that the performance of the SLI group demonstrates a problem in syntax—yet their difference from controls is evident only in passives and *some* types of transitive–intransitive alternation responses but not in all. If the syntactic mechanisms responsible for this production were broken, how is it possible that they function for some types of stimuli? This variation suggests that these mechanisms are present and operative, but the overt realization of their output (“externalization”) might be affected depending on many factors such as the complexity of the task demands (e.g., working memory overload).

Passivization is a classic example of the so-called syntactic deficit. As Penke (2015) notes, most language-impaired

individuals would understand better a canonical SVO structure (e.g., “John kissed Mary”) compared to object clefts (e.g., “It is Mary who John kissed”) or passives (e.g., “Mary was kissed by John”). She notes that language-impaired individuals often misinterpret such structures by interpreting the first NP encountered as AGENT (as in the canonical SVO) instead of THEME. However, this is not a very concrete indication of a syntactic deficit for the following reason: The same mistake (i.e., the strategy to interpret the first NP of a clause as AGENT) is *regularly observed* in control groups that do not have any language impairment whatsoever (Penke, 2015). In other words, the same strategy is employed by healthy neurotypical subjects that have an intact, fully functional syntactic domain. This probably happens because the human parser establishes a threshold for the interpretation of each chunk of input. As noted in Leivada (2015), the strategy that Penke (2015) describes can be connected to the Moses illusion (Reder and Kusbit, 1991), according to which neurotypical individuals are unable to detect distortions in the experimental stimuli such as “How many animals of each kind did Moses take in the Ark?” They might fail to detect the distortion even if they do know that it was Noah and not Moses who built the Ark. This phenomenon has been explained in the literature through recognizing the existence of a *processing threshold* by means of suggesting that a partial-match strategy is operative when the stimuli is processed (Kamas et al., 1996).

Pragmatic cues are very important when the parser establishes this threshold. For example, in relation to the Moses illusion, Moses and Noah are both biblical characters and as such loosely associated in a way that can trick the parser; if Nixon was used instead of Moses, it is much more likely that the distortion would be spotted (Kamas et al., 1996). Observing that all this happens in the case of neurotypical speakers, there is no reason not to capture the problems in passivization in (a)typical speakers in the same uniform way. It has been long noted that reversible passives (e.g., “The boy is being chased by the girl” and “The girl is being chased by the boy”) are more difficult to interpret compared to non-reversible passives which are at least pragmatically odd when reversed (e.g., “The task was carried out by John” and #“John was carried out by the task”), *even in instances of typical language abilities* (Rondal, 2007). Therefore, it comes as no surprise that many atypical populations show a *selective* impairment of passives: Reversible passives are impaired, while non-reversible passives are better preserved (see Caramazza and Miceli, 1991 for aphasia). In this context, it is somewhat expected that in atypical populations that have processing limitations (and many studies attest to this for SLI; see **Table 1**), lower accuracy will be observed in the comprehension of some passives—not because syntax is impaired, but because the partial-match process may operate at an overall lower threshold level perturbing comprehension.

Returning to studies that put forth a syntactic impairment, Marinis and van der Lely (2007) claim that children with SLI show a particular deficit in the computational system that affects syntactic dependencies involving syntactic movement: In contrast to controls, children with SLI showed no priming effect that would indicate a filler–gap dependency. At the same time, their very high performance (ca. 90% accuracy) suggests that they were somehow able to interpret the stimuli correctly. The

priming effect that the results of Marinis and van der Lely (2007) showed at the verb position indicates that an association between two different syntactic positions was indeed established, which in turns means that the ability to form such associations remains operative in SLI populations. This begs the question: Which are then the factors that lead to what many studies describe as impaired or defective syntax?

On the basis of the studies presented in **Table 1**, we suggest that poor memory resources (Montgomery, 2004; Bishop and Donlan, 2005), Theory of Mind deficits (Tsimpli et al., 2017b), and spell-out errors (Lin, 2007; Mastropavlou and Tsimpli, 2011) can explain why a claim for impaired syntax is put forth. For example, Schuele and Dykes (2005) argue in their longitudinal study that certain aspects of syntax may be developed late. They report omissions of infinitival *to*, *wh*-pronouns in clausal interrogative complements, and relative markers. Importantly, this result is cross-linguistically supported (see Mastropavlou and Tsimpli, 2011 for omissions of such functional markers in Standard Modern Greek). Still, one cannot conclude that such omissions occur because these syntactic nodes are broken for two reasons. The first reason is that, even if a functional element is absent, its selectional requirements are fulfilled. The findings of Mastropavlou and Tsimpli (2011) show exactly this pattern:

This leads to a paradoxical situation where the complementizer may be omitted and, hence, not merged in the syntactic position, whereas its selectional restrictions are still operative. This is particularly relevant to the omission cases of *na* which, as mentioned above, is the only complementizer which can introduce tense-dependent verb forms, i.e., the non-past, perfective form. [...] We must, therefore, conclude that even in the case of omissions, children know the selectional properties imposed by C and fail to access or spell-out the required complementizer” (Mastropavlou and Tsimpli, 2011, p. 460).

The second reason boils down to the fact that such omissions are *never* consistent; the markers in question are sometimes produced and sometimes omitted within a single speaker’s productions. If we accept that these omissions are due to a retrieval problem at the level of externalization, we can explain the variation observed across productions as the result of any of the following factors as well as their possible interactions:

- i. The presence or absence of salient pragmatic cues,
- ii. Complexity and task-demand factors (that are related to memory limitations), and
- iii. The (non-)salient morphophonological substance of the omitted markers (in line with the Surface Hypothesis, see Montgomery and Leonard, 1998).

If, however, the locus of impairment is the inability to construct a syntactic representation past a particular node, how is it possible that many times this syntactic representation is constructed and the problematic node surfaces intact? To give an example, if one suggests that the T(ense) node is problematic in Greek SLI, what explains that some affected persons might produce atypical realizations of T at times, while correctly producing T (and nodes past it) other times?



TABLE 1 | SLI studies across different language groups and sets of tasks.

Study	Task(s)	Results	Domain(s) of impairment
<b>(A) OVERVIEW OF STUDIES ON SLI: ENGLISH</b>			
O'Hara and Johnston, 1997	Acting out auditory stimuli with toys	Lower accuracy than controls in acting out the meaning of the stimuli	Processing limitations
Loeb et al., 1998	Verb elicitation and causative alternation	Problems with (i) the production of passive forms and (ii) the transitive-intransitive alternation	Syntax
Montgomery and Leonard, 1998	On-line word recognition and off-line grammaticality judgment	Sensitivity to the presence of high-substance inflection	Morphophonology due to general processing capacity limitation (Surface Hypothesis)
Leonard et al., 2003	Elicited production of past progressive <i>was/were ± ing</i> , present progressive <i>is/are ± ing</i> , third person singular <i>-s</i> in habitual action contexts, modal <i>can</i> in ability/possibility contexts	Affected tense marking	Morphology
Montgomery, 2004	Picture-pointing sentence comprehension and non-word repetition	Poor long non-word repetition and poor comprehension of normal-rate sentences	Phonological working memory
Bishop and Donlan, 2005	Story retelling	Poorer story recall than controls	Difficulties in encoding and remembering meaningful sequences of events
Schuele and Dykes, 2005	Conversational samples	Omissions of (i) infinitival <i>to</i> , (ii) relative markers and pronouns, (iii) <i>wh</i> -pronouns in embedded <i>wh</i> -clauses	Delayed emergence of complex syntax
Lin, 2007	Corpus study	Errors in tense marking and agreement Intact case marking and A-movement	Agreement marking errors as spell-out errors
Marinis and van der Lely, 2007	Cross-modal picture priming	90% accuracy in question comprehension	Syntactic filler-gap dependencies
Marshall et al., 2009	Chunking (receptive task, expressive task), focus (receptive task, expressive task), long-item (discrimination-receptive task, imitation-expressive task)	Failure to disambiguate when prosody is required to interact with syntax/discourse Ability to imitate or discriminate prosodic structures without reference to meaning	Integration of prosody in overall processing
Riches et al., 2010	Sentence repetition	Errors on the more complex object relative clauses	Syntax due to short-term memory limitations
Owen van Hone and Lin, 2011	Conversational and narrative samples	Use of cognitive state verbs, especially high-frequency ones, but reduced combination of low-frequency verbs with complement clauses compared to controls	Lexical knowledge limitations
Claessen and Letiño, 2012	Auditory lexical discrimination, phoneme deletion	Lower quality phonological representations	Phonology
Riches, 2012	Sentence repetition	Errors in sentence repetition and difficulties in delayed repetition	Syntactic representation and maintenance of long-term memory representations in short-term memory
Leonard et al., 2013	Sentence comprehension	Intact syntactic and lexical representations Reduced accuracy in high-demand processing	Processing limitations
Marinis and Saddy, 2013	Off-line picture selection, self-paced listening and picture verification	Difficulties in reanalyzing thematic roles in both actives and passives	Processing limitations
Tomas et al., 2015	Elicited production of past tense <i>-ed</i> , third person singular <i>-s</i> , possessive marker <i>-s</i>	Difficulties with syllabic allomorphs	Morphophonology
<b>(B) OVERVIEW OF STUDIES ON SLI: STANDARD MODERN GREEK</b>			
Claussen and Datalakis, 1999	Spontaneous speech	Subject-verb agreement at chance level due to overproduction of 3rd person intact past tense markings	Morphology
Tsimpli and Stavrakaki, 1999	Spontaneous speech	Omission of direct object and definite articles Mastery of strong pronouns and indefinite articles	Grammar
Kateri et al., 2005	Picture naming	Non-target production of certain phonemes	Phonology
Manika et al., 2010	Elicited production	At ceiling production of direct object clitics and definite articles	—

(Continued)

TABLE 1 | Continued

Study	Task(s)	Results	Domain(s) of Impairment
Mastropavlou, 2010	Sentence completion	Tense marking errors which are less featured on morphophonologically salient forms	Morphology
Stavrakaki and van der Lely, 2010	Elicited production, picture selection	Poorer production and comprehension of object clitics compared to controls	Syntax
Mastropavlou and Tsimpli, 2011	Spontaneous speech	Complementizer omission (for some types of complementizers)	Morphophonology (omissions as spell-out errors)
Laliofi et al., 2016	Non-word repetition, elicited production, off-line grammaticality judgment, on-line self-paced listening	Intact subject-verb agreement, slower reaction times compared to controls, lower performance in the grammaticality judgment task, limited verbal short-term memory	Verbal short-term memory
Tsimpli et al., 2017b	Clitic production, picture-based story retelling	Lower accuracy in first person clitics compared to controls	Theory of Mind deficits
<b>(C) OVERVIEW OF STUDIES ON SLI: STANDARD MODERN GREEK AND CYPRIOT GREEK</b>			
Petinou and Terzi, 2002	Spontaneous speech	Clitic displacement errors	Grammar
Spanoudis et al., 2007	Elicited production	Poor knowledge of mental state verbs, poor inference drawing	Semantics Pragmatics
Kambanaros, 2013, 2014; Kambanaros et al., 2014	Picture naming, connected speech, naming of semantically and phonologically complex verbs	Naming errors Better retrieval of nouns than verbs No difference between noun and verb use in connected speech Greater difficulty naming semantically complex but not phonologically complex verbs	Access to phonological representations
Theodorou and Grohmann, 2015	Elicited production of clitics	Intact clitic production	—
<b>(D) OVERVIEW OF STUDIES ON SLI: SPANISH</b>			
Bedore and Leonard, 2001	Elicited production of present first/third person singular/plural, past first/third person singular/plural, infinitive, definite articles, indefinite articles, direct object clitic pronouns, noun plural inflections, and adjective agreement inflections	Limited use of inflectional morphemes (e.g., agreement markers and direct object clitics)	Morphology
Resitrepo and Gutierrez-Olallen, 2001	Interview, picture description, story retelling	Errors in unstressed definite articles	Morphophonology (Surface Hypothesis)
Katsos et al., 2011	Acceptability judgment task	Difficulty with employing the maxim of quantity	Pragmatics
Gimstead et al., 2013	Spontaneous speech, repetition task	Affected tense markings	Morphology
Andreu et al., 2016	Sentence comprehension	Intact accessing of the semantic information of verbs	—
Gavarró and Cantú-Sánchez, 2016	Vocabulary retrieval, non-word repetition, elicitation of morphology and syntax	Intact vocabulary, non-word repetition, determiners, sentence repetition, subject and object relatives, preposition, nominal agreement, and pluralization	—
<b>(E) OVERVIEW OF STUDIES ON SLI: SPANISH AND CATALAN</b>			
Aguilar-Mediavilla and Serra-Raventós, 2006	Interview	Delay in the acquisition of segments, complex syllabic structures, and differences compared to controls in the production of various segments	Phonology
Sanz-Torrent et al., 2008	Spontaneous speech	Verb inflection	Morphology (Surface Hypothesis)
Andreu et al., 2013	Picture description	Omissions of obligatory arguments (themes) as complexity increased	Limited semantic representations
Aguilar-Mediavilla et al., 2014	Navarra Oral Language Test-Revised (PLON-R; Prueba del Lenguaje Oral de Navarra-Revisada)	Reduced phonological awareness and memory, letter identification, and semantic comprehension	Processing limitations Phonology
Buil-Legaz et al., 2016	Referential communication task	Less informative messages compared to controls	Semantics Processing limitations Pragmatics
Gavarró and Lite, 2015 [for Catalan]	Truth value judgment task	Intact semantics of quantifiers	Processing limitations

Having analyzed 682 linguistic profiles of children with SLI, we observe that the loci of impairment are related to externalization: morphophonology and pragmatics. Variation is attested in *some* parts of the language faculty and often appears in the form of omissions that occur due to retrieval/spell-out errors (Lin, 2007; Mastropavlou and Tsimpli, 2011), delayed mastery of phonology and failure to integrate related cues in overall processing (Kateri et al., 2005; Marshall et al., 2009), and processing, memory, and pragmatic limitations (Montgomery, 2004; Bishop and Donlan, 2005; Tsimpli et al., 2017b).

## Comparing SLI with Other Disorders: ASD and DS

Pragmatic difficulties and morphophonological omissions are not restricted to SLI. The literature on ASD and DS has repeatedly highlighted the existence of such features in the linguistic profiles of these populations.

### Autism Spectrum Disorder (ASD)

Starting off with ASD, the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013 p. 809) defines it as “characterized by deficits in two core domains: (1) deficits in social communication and social interaction and (2) restricted repetitive patterns of behavior, interests, and activities.”

The existence of repetitive patterns is one of the key characteristics of ASD language. Kanner (1943) was the first to describe instances of “parroting,” echolalia, and atypical use of personal pronouns that involved pronoun repetition in the autistic subjects of his study. Much subsequent work has focused on pronoun reversals (i.e., use of “you” instead of “I”) in ASD and many explanations have been offered for this phenomenon, including that of echolalia (Kanner’s original explanation), impaired discourse understanding, and impaired Theory of Mind (see Brehme, 2014 for a review). Some studies have described such reversals as grammatical errors, a description that may imply a deficient language module with impaired syntax (Bartolucci and Albers, 1974; Belkadi, 2006; Wittke et al., 2017).

Apart from pronoun reversals, some ASD linguistic profiles feature other types of grammatical errors (mainly in verbal, nominal, and pronominal morphology) in around 27–28% of their utterances, a result comparable to the frequency of grammatical errors in SLI (Wittke et al., 2017). Morphology stands out as a vulnerable domain once more, and so do pragmatic abilities in ASD (Lord and Paul, 1997; Volden et al., 2009; Marinis et al., 2013).

Syntax in ASD has received mixed descriptions. On the one hand, Bartolucci and Albers (1974, p. 131) begin their study of tense marking in autism by postulating a syntactic problem: “Certain characteristics of the syntactic structures of the language of autistic children, such as their lack of mastery of pronominalization, have been described.” On the other hand, many reviews have concluded that ASD syntax is not deficient, since many syntactic dependencies remain intact especially in high-functioning individuals, but merely follow typical development at a slower rate (Tager-Flusberg, 1981; Perovic and Janke, 2013). Other studies revealed *subtle*

difficulties in some syntactic measures, regardless of language development history (Durrleman et al., 2015). Looking at the relevant results, the notion of a processing threshold that was earlier discussed in relation to SLI becomes relevant for ASD too. For instance, the ASD group in Durrleman et al. (2015) obtained lower scores in the comprehension of object relative clauses compared to subject relative clauses. This asymmetry could boil down to the non-canonical word order derived by the fronted object in object relative clauses. In other words, this additional layer of complexity could be responsible for the subject–object asymmetry that is observed in the comprehension of relative clauses not only in ASD and SLI, but also in *neurotypical* populations, with subject relatives usually being easier to process (see Carreiras et al., 2010 for a review and a counterexample).

Returning to the atypical use of pronouns, Kanner (1943) and many subsequent studies indeed offer data that involve pronoun reversals. However, they also offer examples (of the same children, at the same stage of development) that show target use of pronouns (Leivada, 2015). If these pronoun reversals were the outcome of broken syntax, how is it possible that the target performance emerges at times? Put differently, if the locus of the deficiency is to be found in the innermost component of language (i.e., syntax), what makes possible the externalization of the target pattern often in a consistent fashion?

Interestingly, use of pronouns is not always atypical in ASD. Some studies have revealed high accuracy in the comprehension of different types of pronouns including strong pronouns, clitics, and reflexives (Terzi et al., 2012, 2014 for Standard Modern Greek). In these studies, the lowest performance was found in the clitics condition (mean correct: 88.3%) for which the most frequent error was theta-role reversals. Is this an indication of deficient syntax? As Terzi et al. (2012, 2014) show, these children had problems with producing clitic pronouns, so it is not clear whether their low performance in the clitics condition is the result of a problem in syntactic binding or the particular grammar of clitics. Terzi et al. (2014) carried out a follow-up study that aimed to clarify this issue. The results showed that the ASD group produced a high number of clitics, yet a lower one compared to the control group (87.39% correct vs. 97.74% correct, respectively), thus favoring the scenario that renders clitics and not binding responsible for the lower performance in the clitic condition of the task.

This lower performance of the ASD group in the clitics condition is compatible with the idea put forth in the present work that loci of impairment are confined to certain parts of the language faculty. We have argued that morphology and pragmatics are shown to be vulnerable across pathologies, languages, and elicitation tasks. Clitics are markers of morphological agreement, licensed under specific pragmatic conditions, and children with ASD have troubles in ascertaining what is prominent/salient in the discourse (Terzi et al., 2016). In a subsequent study that involved narratives instead of a highly structured elicitation task, Terzi et al. (2017) found that the same group of ASD children did produce clitics, a fact that highlights the importance of the tool used to elicit data. According to Terzi et al. (2017, p. 648), ASD children “had full control of the discourse by contrast to the structured experiments, the nature of

which was such that they had to take into account the discourse representation provided by the experimenter in each condition and trial.”

In this context, it could well be the case that pronoun reversals in ASD do not stem from impaired language/syntax. Studies of deaf children with autism provided below lend support for this hypothesis. It has been suggested in the relevant literature that what seems to be at stake in ASD is a *less secure anchorage in self-experience* (Lee et al., 1994). Shield and Meier’s (2014) experiment is instrumental in evaluating this hypothesis. They showed deaf autistic and deaf typically developing children a picture of themselves and a picture of the experimenter. Upon seeing a picture of themselves and being asked “Who is this?” the children with ASD either signed the pronoun “me” pointing to themselves or produced their name sign or finger-spelled their English name. In other words, they were successful both in identifying themselves and in using the correct pronoun, whenever a pronoun was used. The same strategies were employed by the typically developing group. What differentiated the two groups is not self-identification *per se* or the linguistic strategy through which self-identification was achieved, but the fact that the typically developing children “reacted with a smile or laugh and an emphatic point at his/her own body. The children with ASD had no such emotional reaction.” (Shield and Meier, 2014, p. 412). As the authors note in their discussion of these findings, forming a sense of me-ness is a key component of social behaviors such as empathy.

This less secure anchorage in me-ness can be manifested in ways that have nothing to do with the use of pronouns, thereby suggesting that the pronoun-reversal problem is not linguistic or syntactic as such (Leivada, 2015). A crucial piece of evidence that leads to this conclusion comes from studies of palm orientation during signing by deaf children with ASD. Shield and Meier (2012) found that native signers of American Sign Language with ASD showed a tendency to reverse palm orientation on signs specified for inward/outward orientation, whereas such errors were absent from the production of their typically developing peers. Observing this atypical anchorage in selfhood, one can suggest that their linguistic/grammatical counterparts (i.e., pronoun reversals) reflect not a syntactic problem but rather a more general cognitive problem that may acquire a linguistic dress (Leivada, 2015). If this observation is on the right track, syntax seems to be unimpaired in ASD, whereas other domains of language such as morphophonology (Kanner, 1943) and pragmatics (Terzi et al., 2014) stand out as particularly susceptible to impairment.

### Down Syndrome (DS)

DS is the result of a genetic abnormality most often caused from the presence of a third chromosome 21. One of the characteristics of this syndrome is atypical cognitive development. When it comes to language, our review of studies on DS suggest it is somewhat challenging when one pursues a claim of preserved syntax (as some studies have identified syntactic deficits in the profile of their subjects; e.g., Perovic, 2001).

One domain of language that has been argued to be atypical in DS is syntactic binding. Binding Theory regulates the distribution

of referentially dependent elements such as anaphors and pronouns (Chomsky, 1981). Binding Principle A requires that the anaphor is locally bound by an antecedent within the same clause/domain (e.g., Mary<sub>i</sub> criticized herself<sub>i/\*j</sub>). Principle B requires that the antecedent of a pronoun be not in the same clause/domain as the pronoun (e.g., Bill<sub>j</sub> said that John<sub>i</sub> criticized him<sub>j/\*i</sub>). Principle C prohibits a referential expression from being c-commanded by a coindexed element (e.g., He<sub>i</sub>/Bill<sub>j</sub> criticized John<sub>j/\*i</sub>).

Investigating the comprehension abilities of English-speaking adolescents with DS using a truth-value judgment task, Perovic (2001) found at ceiling performance on the “name-pronoun” condition (e.g., “Is Snow White washing her?”) and high performance ( $\geq 75\%$ ) for the “quantifier-pronoun” condition (e.g., “Is every bear washing him?”). This suggests that whatever the syntactic deficit amounts to, it is not Principle B. The conditions “name-reflexive” (e.g., “Is Snow White washing herself?”) and “quantifier-reflexive” (e.g., “Is every bear washing himself?”) elicited mixed responses with the percentage of correct answers ranging from 12.50 to 100% correct.

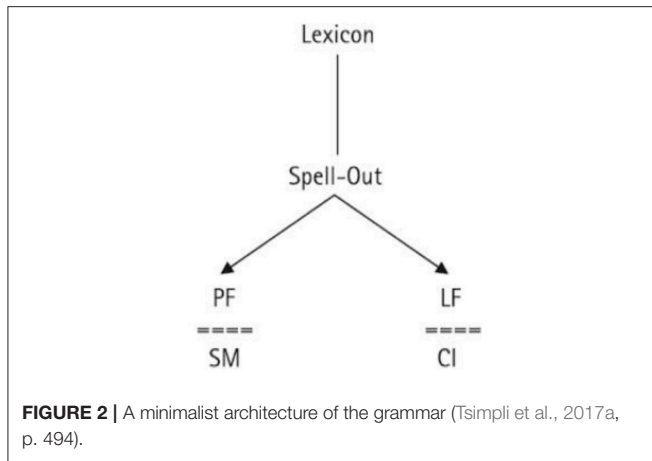
Is Principle A an example of deficient syntax in DS? The answer must be negative for a number of reasons (see Leivada, 2015 for more extensive discussion). First, would be a non-trivial task to explain why individuals with a deficient syntax would face difficulties with one binding principle but not with another, given that all binding principles require the same underlying grammatical knowledge (Perovic, 2001). Second, the results did not show a unanimous pattern of Principle A violations. The average number of correct responses on the “name-reflexive” condition was above chance (56.56% correct). In turn, the average number of correct responses on the “quantifier-reflexive” condition was below chance (35.94% correct), but as Perovic (2001) noted, two participants showed very poor performance even on the control condition that involved quantified NPs and no anaphors. It is then possible that these participants had issues with quantification generally, which resulted in errors on some of the tested conditions.

Tsakiridou (2006) and Christodoulou (2011) focused on Standard Modern Greek and Cypriot Greek DS grammars, respectively. Both showed that the deviations noted in the DS linguistic profile were related to morphophonology: non-target morphological markings (Tsakiridou, 2006) and phonetically or morphophonologically conditioned differences (Christodoulou, 2011). Pragmatics in DS is also atypical. Challenges may include initiation of topics and communicative repairs and aspects of narratives (Martin et al., 2009).

The overall picture that emerges with respect to the linguistic phenotype of DS is one that supports the claim that the aspects of language which appear to be atypical are related to specific parts of the language faculty: morphophonology and pragmatics.

### The Locus Preservation Hypothesis

Having reviewed the literature on three developmental disorders, the first observation is that certain morphological markers reveal themselves as highly susceptible to impairment (e.g., agreement markers and clitics). Second, syntax appears to be preserved. Undoubtedly, some studies have identified problems in the



comprehension or production of complex syntactic structures across disorders (see **Table 1** for SLI). Yet, when considering the general processing limitations that are arguably present in the pathologies discussed in the present work (even though, unfortunately, not fully or equally measured in all studies), we are facing a classical chicken and egg question (Bishop and Donlan, 2005): Do syntactic limitations lead to conceptual and memory limitations or do conceptual and memory limitations result in an atypical syntax?

We have argued that poor memory resources (Montgomery, 2004; Bishop and Donlan, 2005), Theory of Mind deficits (Tsimplici et al., 2017b), and retrieval/spell-out errors (Lin, 2007; Mastropavlou and Tsimplici, 2011) can explain why a claim for impaired syntax is put forth at times. Observing how “linguistic” deficits such as the incorrect use of anaphors in ASD can derive from a general cognitive problem in establishing meanness in relation to the outer world, we tentatively conclude that atypical cognitive abilities (i.e., processing impairments, memory limitations; see **Table 1**) may result in what looks as an atypical syntax. The latter is manifested mainly through omissions, and recall that it would be wrong to conclude that such omissions occur because the related syntactic nodes or operations are broken. The selectional requirements of omitted functional elements may still be operative and satisfied (Mastropavlou and Tsimplici, 2011). Therefore, it makes more sense to describe such omissions as spell-out errors related to the externalization component of language.

Looking at the distribution of impaired and preserved markers/levels of linguistic analysis, variation across pathologies can be formally captured within the *Locus Preservation Hypothesis* (see also Leivada, 2015 for an earlier formulation based on Greek data only):

#### (1) *Locus Preservation Hypothesis*

Syntactic operations are preserved and impenetrable to variation across developmental pathologies.

Assuming a widely accepted architecture of the grammar as the one shown in **Figure 2**, the *Locus Preservation Hypothesis* holds that the computational part of the human language

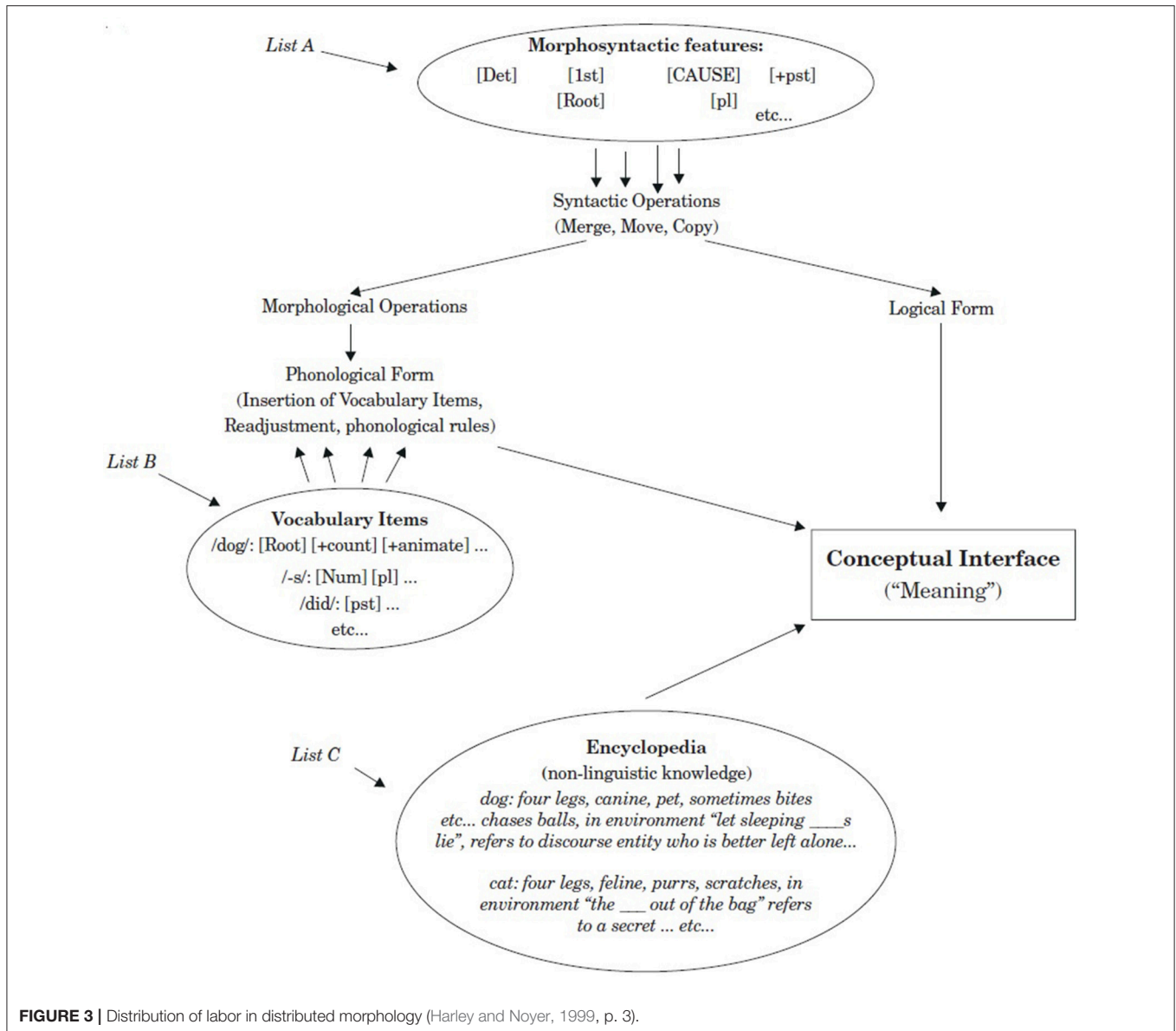
faculty is invariably preserved, with the operations (internal and external) Merge and Agree applying in an intact manner all the way to constructing the internal interface levels of Logical Form (LF) and Phonetic/Phonological Form (PF).

The purported pragmatic deficiencies (Katsos et al., 2011) arise post-syntactically, where the conceptual-intentional system (CI) is accessed along with pragmatic information and encyclopedic/world knowledge. Likewise, the externalization difficulty observed in language production tasks and spontaneous speech (Mastropavlou and Tsimplici, 2011) is relevant at the other interface, the articulatory-perceptual or sensory-motor system (SM). The fact that bound morphophonological building blocks are often misused (Bedore and Leonard, 2001) suggests the need for a finer distinction of the “Lexicon” than what the architecture in **Figure 2** allows.

Mapping the *Locus Preservation Hypothesis* to the distribution of labor put forth in *Distributed Morphology*, it seems that the first set of operations in the transition from List A to List B are resilient to impairment across atypical cognitive phenotypes. In contrast, morphophonological operations and encyclopedic knowledge are consistently susceptible to impairment across atypical cognitive phenotypes (Leivada, 2015). The results that led to this conclusion come from three developmental disorders (SLI, ASD, DS), but there are reasons to believe that this conclusion would hold even when one examines the linguistic profile of acquired pathologies such as aphasia; a topic to be pursued in future work on the *Locus Preservation Hypothesis*. A more detailed model is provided in **Figure 3**.

Overall, based on our review of different research studies, not all pathologies show the same impaired markers—but the same markers are consistently impaired across pathologies. The *Locus Preservation Hypothesis* is thus pathology-independent and can be used to support cross-linguistic findings.

The important question is why syntactic operations are better preserved in a consistent way across disorders with different genetic etiology. One explanation is that the phenotypic overlaps that we identified are in fact surface reflections of more deeply rooted overlaps at the connectome or even the oscillome (Benítez-Burraco and Murphy, 2016). Observing that the hierarchy of brain oscillations has remained remarkably preserved during mammalian evolution (Buzsáki et al., 2013), Benítez-Burraco and Murphy (2016) suggest that language deficits in various cognitive disorders can be traced back to a brain syntax network. In this context, it can be argued that syntax is preserved because it is implemented through a network that is less novel in evolutionary terms, hence more resilient to impairment. Less resilient networks underlie cognitive capacities more recently evolved in phylogenetic terms, whereby selective pressures have not yet given rise to the development of robust compensatory mechanisms (Toro et al., 2010; Murphy and Benítez-Burraco, 2016). This claim grants support to another hypothesis recently explored in the language evolution literature: the “syntax-before-phonology” hypothesis. Based on a review of linguistic calls across species, Collier et al. (2014) argue that syntax, which is universally present in



**FIGURE 3** | Distribution of labor in distributed morphology (Harley and Noyer, 1999, p. 3).

all languages, possibly evolved before phonology, since many systems of communication in other species have the former but not the latter. The Locus Preservation Hypothesis suggests that phonology is less resilient in stark contrast to syntax—a finding that is in line with what the ethological record reveals (Collier et al., 2014).

## CONCLUSIONS

The present work has put forth a novel hypothesis: the Locus Preservation Hypothesis, in order to capture the distribution of what are considered atypical linguistic markers across different languages and pathologies. It has been argued that syntactic operations are resilient to impairment across developmental disorders; in contrast, morphophonology and pragmatics are consistently impaired. This conclusion stands in agreement with

a long line of literature that discusses overlaps in the behavioral profile of populations with different pathologies, both acquired and developmental (Phillips, 2005; Glisky, 2007; Kambanaros and van Steenbrugge, 2013; Benítez-Burraco and Boeckx, 2014; Leivada, 2014, 2015; Kambanaros and Grohmann, 2015; Tsimpli et al., 2017a).

The Locus Preservation Hypothesis can gain more support by expanding the range of languages and pathologies that are examined. Once this is done, the following question to be explored in detail is why syntax would be preserved. One explanation we contemplated in the present work relates to the possibility of an underlying uniform etiology across the reviewed disorders. This uniformity can be traced back to brain network organization (van den Heuvel and Sporns, 2013; Crossley et al., 2014; Benítez-Burraco and Murphy, 2016). Addressing the parallels that can be observed across different

levels of representation (phenome, connectome, dynome, and oscillome) from a phylogenetic perspective, we have established a connection between the hypothesis put forth in the present work and the “syntax-before-phonology” hypothesis of Collier et al. (2014): Syntax is better preserved because it evolved before other domains of language (e.g., morphology and phonology). Therefore, syntax had more adaptation time for the development of compensatory mechanisms, unlike more recently evolved cognitive/linguistic capacities. Future research on the Locus Preservation Hypothesis will elaborate on the syntax-first hypothesis and flesh out the connections between the observed overlap at the phenotypic level and its roots in deeper levels of representation.

## REFERENCES

- Aguilar-Mediavilla, E., Buil-Legaz, L., Pérez-Castelló, J. A., Rigo-Carratalà, E., and Adrover-Roig, D. (2014). Early preschool processing abilities predict subsequent reading outcomes in bilingual Spanish–Catalan children with specific language impairment (SLI). *J. Commun. Disord.* 50, 19–35. doi: 10.1016/j.jcomdis.2014.03.003
- Aguilar-Mediavilla, E., and Serra-Raventós, M. (2006). Phonological profile of Spanish–Catalan children with specific language impairment at age 4: are there any changes over time? *Folia Phoniatr. Logoped.* 58, 400–414. doi: 10.1159/000095001
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders—(DSM-V)*. Washington, DC: American Psychiatric Association.
- Anderson, J. S., Nielsen, J. A., Ferguson, M. A., Burbach, M. C., Cox, E. T., Dai, L., et al. (2013). Abnormal brain synchrony in Down Syndrome. *Neuroimage* 24, 703–715. doi: 10.1016/j.neuroimage.2013.05.006
- Andreu, L., Sanz-Torrent, M., Olmos, J. G., and MacWhinney, B. (2013). The formulation of argument structure in SLI: an eye-movement study. *Clin. Linguist. Phonet.* 27, 111–133. doi: 10.1016/j.clinphon.2012.07.003
- Andreu, L., Sanz-Torrent, M., and Rodríguez-Ferreiro, J. (2016). Do children with SLI use verbs to predict arguments and adjuncts: evidence from eye movements during listening. *Front. Psychol.* 6:1917. doi: 10.3389/fpsyg.2015.01917
- Bartha-Doering, L., Regele, S., Deuster, D., Seidl, R., Bogdanova, N., Röpke, A., et al. (2016). Intrafamilial phenotypic variability of specific language impairment. *Brain Lang.* 159, 102–108. doi: 10.1016/j.bandl.2016.06.005
- Bartolucci, G., and Albers, R. J. (1974). Deictic categories in the language of autistic children. *J. Autism Child. Schizophr.* 4, 131–141. doi: 10.1007/BF02105366
- Bedore, L. M., and Leonard, L. B. (2001). Grammatical morphology deficits in Spanish-speaking children with specific language impairment. *J. Speech Lang. Hear. Res.* 44, 905–924. doi: 10.1044/1092-4388(2001)072
- Belkadi, A. (2006). Language impairments in autism: evidence against mind-blindness. *SOAS Work. Papers Ling.* 14, 3–13.
- Benítez-Burraco, A., and Boeckx, C. (2014). Universal Grammar and biological variation: an EvoDevo agenda for comparative biolinguistics. *Biol. Theory* 9, 122–134. doi: 10.1007/s13752-014-0164-0
- Benítez-Burraco, A., and Murphy, E. (2016). The oscillopathic nature of language deficits in autism: from genes to language evolution. *Front. Hum. Neurosci.* 10:120. doi: 10.3389/fnhum.2016.00120
- Bishop, D. V. M. (2004). “Specific Language Impairment: diagnostic dilemmas,” in *Classification of Developmental Language Disorders*, eds L. Verhoeven and H. van Balkom (Mahwah, NJ: Lawrence Erlbaum), 309–326.
- Bishop, D. V. (2014a). Ten questions about terminology for children with unexplained language problems. *Int. J. Lang. Commun. Disord.* 49, 381–415. doi: 10.1111/1460-6984.12101
- Bishop, D. V. (2014b). Problems with tense marking in children with specific language impairment: not how but when. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20120401. doi: 10.1098/rstb.2012.0401
- Bishop, D. V. M., and Donlan, C. (2005). The role of syntax in encoding and recall of pictorial narratives: evidence from specific language impairment. *Br. J. Dev. Psychol.* 23, 25–46. doi: 10.1348/026151004X20685
- Bobaljik, J. D. (2017). Distributed morphology. *Oxford Res. Encyclop. Linguist.* doi: 10.1093/acrefore/9780199384655.013.131
- Brehme, D. F. (2014). Perspectives on personal pronoun reversal in children with ASD: a critical review. *J. Euro. Psychol. Stud.* 5, 31–37. doi: 10.5334/jeps.br
- Buil-Legaz, L., Pérez-Castelló, J. A., Adrover-Roig, D., and Aguilar-Mediavilla, E. (2016). Referential communication in bilingual Catalan–Spanish children with specific language impairment. *Anuar. Psicol.* 46, 31–40. doi: 10.1016/j.anpsic.2016.06.004
- Buzsáki, G., Logothetis, N., and Singer, W. (2013). Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron* 80, 751–764. doi: 10.1016/j.neuron.2013.10.002
- Caramazza, A., and Miceli, G. (1991). Selective impairment of thematic role assignment in sentence processing. *Brain Lang.* 41, 402–436. doi: 10.1016/0093-934X(91)90164-V
- Carreiras, M., Duñabeitia, J. A., Vergara, M., de la Cruz-Pavía, I., and Laka, I. (2010). Subject relative clauses are not universally easier to process: evidence from Basque. *Cognition* 115, 79–92. doi: 10.1016/j.cognition.2009.11.012
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N. (2001). “Derivation by phase,” in *Ken Hale: A Life in Language*, ed M. Kenstowicz (Cambridge, MA: MIT Press), 1–52.
- Christodoulou, C. (2011). *Cypriot Greek Down Syndrome: Their Grammar and Its Interfaces*. Doctoral Dissertation, The University of British Columbia.
- Claessens, M., and Leitão, S. (2012). Phonological representations in children with SLI. *Child Lang. Teach. Ther.* 28, 211–223. doi: 10.1177/0265659012436851
- Clahsen, H., and Dalalakis, J. (1999). Tense and agreement in Greek SLI: a case study. *Essex Res. Rep. Linguist.* 24, 1–25.
- Collier, K., Bickel, B., van Schaik, C. P., Manser, M. B., and Townsend, S. W. (2014). Language evolution: syntax before phonology? *Proc. R. Soc. B* 281:20140263. doi: 10.1098/rspb.2014.0263
- Crossley, N. A., Mechelli, A., Scott, J., Carletti, F., Fox, P. T., McGuire, P., et al. (2014). The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* 137, 2382–2395. doi: 10.1093/brain/awu132
- Dodwell, K., and Bavin, E. L. (2008). Children with specific language impairment: an investigation of their narratives and memory. *Int. J. Lang. Commun. Disord.* 43, 201–218. doi: 10.1080/13682820701366147
- Durrleman, S., and Delage, H. (2016). Autism spectrum disorder and specific language impairment: overlaps in syntactic profiles. *Lang. Acquisit.* 23, 361–386. doi: 10.1080/10489223.2016.1179741
- Durrleman, S., Hippolyte, L., Zufferey, S., Iglesias, K., and Hadjikhani, N. (2015). Complex syntax in autism spectrum disorders: a study of relative clauses. *Int. J. Lang. Commun. Disord.* 50, 260–267. doi: 10.1111/1460-6984.12130

## AUTHOR CONTRIBUTIONS

All authors participated in the analysis. EL drafted the manuscript. MK and KG reviewed and revised the manuscript.

## ACKNOWLEDGMENTS

We thank the two reviewers and Nikoleta Christou for helpful comments. We also thank the audiences of the conferences where this work was presented in Vienna, Patras, and Barcelona. The publication charges for this article have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

- Dykens, E. M., Lee, E., and Roof, E. (2011). Prader-Willi syndrome and autism spectrum disorders: an evolving story. *J. Neurodev. Disord.* 3, 225–237. doi: 10.1007/s11689-011-9092-5
- Eadie, P. A., Fey, M. E., Douglas, J. M., and Parsons, C. L. (2002). Profiles of grammatical morphology and sentence imitation in children with specific language impairment and Down syndrome. *J. Speech Lang. Hear. Res.* 45, 720–732. doi: 10.1044/1092-4388(2002/058)
- Engel de Abreu, P. M., Cruz-Santos, A., and Puglisi, M. L. (2014). Specific language impairment in language-minority children from low-income families. *Int. J. Lang. Commun. Disord.* 49, 736–747. doi: 10.1111/1460-6984.12107
- Ferreira, F. (2005). Psycholinguistics, formal grammars, and cognitive science. *Linguist. Rev.* 22, 365–380. doi: 10.1515/tlir.2005.22.2-4.365
- Fowler, A. E. (1995). “Linguistic variability in persons with down syndrome: research and implications,” in *Down Syndrome: Living and Learning in the Community*, eds L. Nadel and D. Rosenthal (New York, NY: Wiley-Liss), 121–131.
- Friedmann, N., and Novogrodsky, R. (2008). “Subtypes of SLI: SySLI, PhoSLI, LeSLI, and PraSLI,” in *Language Acquisition and Development: Proceedings of GALA 2007*, eds A. Gavarró and M. J. Freitas (Newcastle: Cambridge Scholars Publishing), 205–217.
- Gathercole, S., and Baddeley, A. (1990). Phonological memory deficits in language disordered children: is there a causal connection? *J. Mem. Lang.* 29, 336–360. doi: 10.1016/0749-596X(90)90004-J
- Gavarró, A., and Cantú-Sánchez, M. (2016). “A study on an alleged case of Spanish SLI and the founder effect,” in *Biolinguistic Investigations on the Language Faculty*, ed A. M. Di Sciullo (Amsterdam: John Benjamins), 169–193.
- Gavarró, A., and Lite, A. (2015). “Universal quantification in Catalan SLI,” in *Specific Language Impairment: Current Trends in Research*, ed S. Stavrakaki (Amsterdam: John Benjamins), 191–213.
- Geschwind, D. H. (2011). Genetics of autism spectrum disorders. *Trends Cogn. Sci.* 15, 409–416. doi: 10.1016/j.tics.2011.07.003
- Glisky, E. L. (2007). “Changes in cognitive function in human aging,” in *Brain Aging: Models, Methods, and Mechanisms*, ed D. R. Riddle (Boca Raton, FL: CRC Press), 4–20.
- Grinstead, J., Baron, A., Vega-Mendoza, M., De la Mora, J., Cantú-Sánchez, M., and Flores, B. (2013). Tense marking and spontaneous speech measures in Spanish specific language impairment: a discriminant function analysis. *J. Speech Lang. Hear. Res.* 56, 352–363. doi: 10.1044/1092-4388(2012/11-0289)
- Hagoort, P. (2014). “Linguistics quo vadis? An outsider perspective,” in *Talk Presented at SLE 2014: 47th Annual Meeting of the Societas Linguistica Europaea* (Poznan: Adam Mickiewicz University). Available online at: <http://thescienceoflanguage.com/2015/02/16/linguistics-quo-vadis-an-outsider-perspective>
- Halle, M., and Marantz, A. (1993). “Distributed morphology and the pieces of inflection,” in *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, eds K. Hale and S. J. Keyser (Cambridge, MA: MIT Press), 111–176.
- Harley, H., and Noyer, R. (1999). Distributed morphology. *Glott. Int.* 4, 3–9.
- Kamas, E. N., Reeder, L. M., and Ayers, M. S. (1996). Partial matching in the Moses illusion: response bias not sensitivity. *Mem. Cogn.* 24, 687–699. doi: 10.3758/BF03201094
- Kampanaros, M. (2013). Does verb type affect action naming in specific language impairment (SLI)? Evidence from instrumentality and name relation. *J. Neurolinguist.* 26, 160–177. doi: 10.1016/j.jneuroling.2012.07.003
- Kampanaros, M. (2014). Context effects on verb production in specific language impairment (SLI): confrontation naming versus connected speech. *Clin. Linguist. Phonet.* 28, 826–843. doi: 10.3109/02699206.2014.911962
- Kampanaros, M., and Grohmann, K. K. (2015). Grammatical class effects across impaired child and adult populations. *Front. Psychol. Lang. Sci.* 6:1670. doi: 10.3389/fpsyg.2015.01670
- Kampanaros, M., Grohmann, K. K., Michaelides, M., and Theodorou, E. (2014). On the nature of verb–noun dissociations in bilingual SLI: a psycholinguistic perspective from Greek. *Bilingualism* 17, 169–188. doi: 10.1017/S1366728913000035
- Kampanaros, M., and van Steenbrugge, W. (2013). Lexical retrieval deficits in anomia aphasia and specific language impairment (SLI): more similar than different? Grammatical class and context effects. *Linguist. Variat.* 13, 237–256. doi: 10.1075/lv.13.2.05kam
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nerv. Child* 2, 217–250.
- Kateri, L., Revithiadou, A., and Varlokosta, S. (2005). “Multiple partially ordered grammars: a case study from Greek SLI,” in *Developmental Paths in Phonological Acquisition*, eds M. Tzakosta, C. Levelt and J. van de Weijer (Leiden: Leiden Centre for Linguistics), 13–38.
- Katsos, N., Roqueta, C. A., Estevan, R. A., and Cummins, C. (2011). Are children with specific language impairment competent with the pragmatics and logic of quantification? *Cognition* 119, 43–57. doi: 10.1016/j.cognition.2010.12.004
- Lalioti, M., Stavrakaki, S., Manouilidou, C., and Talli, I. (2016). Subject–verb agreement and verbal short-term memory: a perspective from Greek children with specific language impairment. *First Language* 36, 279–294. doi: 10.1177/0142723716648844
- Lecavalier, L. (2014). “Phenotypic variability in autism spectrum disorder: clinical considerations,” in *Hand book of Autism and Anxiety*, eds T. E. Davis III, S. W. White and T. H. Ollendick (Cham: Springer International Publishing), 15–29.
- Lee, A., Hobson, R. P., and Chiat, S. (1994). I, you, me, and autism: an experimental study. *J. Autism Dev. Disord.* 24, 155–176. doi: 10.1007/BF02172094
- Leivada, E. (2014). From comparative linguistics to comparative (bio)linguistics: reflections on variation. *Biolinguistics* 8, 53–66.
- Leivada, E. (2015). *The Nature and Limits of Variation across Languages and Pathologies*. Doctoral Dissertation, Universitat de Barcelona.
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York, NY: John Wiley and Sons.
- Leonard, L. B. (1998). *Children with Specific Language Impairment*. Cambridge, MA: MIT Press.
- Leonard, L. B. (2014). Specific language impairment across languages. *Child Dev. Perspect.* 8, 1–5. doi: 10.1111/cdep.12053
- Leonard, L. B., Deevy, P., Fey, M. E., and Bredin-Oja, S. L. (2013). Sentence comprehension in specific language impairment: a task designed to distinguish between cognitive capacity and syntactic complexity. *J. Speech Lang. Hear. Res.* 56, 577–589. doi: 10.1044/1092-4388(2012/11-0254)
- Leonard, L. B., Deevy, P., Miller, C. A., Charest, M., Kurtz, R., and Rauf, L. (2003). The use of grammatical morphemes reflecting aspect and modality by children with specific language impairment. *J. Child Lang.* 30, 769–795. doi: 10.1017/S0305000903005816
- Leyfer, O. T., Tager-Flusberg, H., Dowd, M., Tomblin, J. B., and Folstein, S. E. (2008). Overlap between autism and specific language impairment: comparison of autism diagnostic interview and autism diagnostic observation schedule scores. *Autism Res.* 1, 284–296. doi: 10.1002/aur.43
- Lin, Y.-A. (2007). “On grammatical errors in English SLI children: a corpus-based study,” in *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA)*, eds A. Belikova, L. Meroni and M. Umeda (Somerville, MA: Cascadia Proceedings Project), 245–252.
- Loeb, D. F., Pye, C., Richardson, L. Z., and Redmond, S. (1998). Causative alternations of children with specific language impairment. *J. Speech Lang. Hear. Res.* 41, 1103–1114. doi: 10.1044/jslhr.4105.110
- Lord, C., and Paul, R. (1997). “Language and communication in autism,” in *Hand book of Autism and Pervasive Developmental Disorders, 2nd Edn*, eds F. Volkmar and D. Cohen (New York, NY: John Wiley and Sons), 195–225.
- Manika, S., Varlokosta, S., and Wexler, K. (2010). “The lack of omission of clitics in Greek children with SLI: an experimental study,” in *Proceedings of the 35th Annual Boston University Conference on Language Development, Vol. 35*, 427–439.
- Marantz, A. (2005). Generative linguistics within the cognitive neuroscience of language. *Linguist. Rev.* 22, 429–445. doi: 10.1515/tlir.2005.22.2-4.429
- Marinis, T., and Saddy, D. (2013). Parsing the passive: comparing children with specific language impairment to sequential bilingual children. *Lang. Acquisit.* 20, 155–179. doi: 10.1080/10489223.2013.766743
- Marinis, T., Terzi, A., Kotsopoulou, A., and Francis, K. (2013). Pragmatic abilities of high-functioning Greek-speaking children with autism. *Psychologia* 20, 321–337.



- Marinis, T., and van der Lely, H. K. (2007). On-line processing of wh-questions in children with G-SLI and typically developing children. *Int. J. Lang. Commun. Disord.* 42, 557–582. doi: 10.1080/13682820601058190
- Marshall, C. R., Harcourt-Brown, S., Ramus, F., and van der Lely, H. K. J. (2009). The link between prosody and language skills in children with specific language impairment (SLI) and/or dyslexia. *Int. J. Lang. Commun. Disord.* 44, 466–488. doi: 10.1080/13682820802591643
- Martin, G. E., Klusek, J., Estigarribia, B., and Roberts, J. E. (2009). Language characteristics of individuals with Down syndrome. *Top. Lang. Disord.* 29, 112–132.
- Mastropavlou, M. (2010). Morphophonological salience as a compensatory means for deficits in the acquisition of past tense in SLI. *J. Commun. Disord.* 43, 175–198. doi: 10.1016/j.jcomdis.2009.12.005
- Mastropavlou, M., and Tsimpli, I. M. (2011). Complementizers and subordination in typical language acquisition and SLI. *Lingua* 121, 442–462. doi: 10.1016/j.lingua.2010.10.009
- Montgomery, J. W. (2004). Sentence comprehension in children with specific language impairment: effects of input rate and phonological working memory. *Int. J. Lang. Commun. Disord.* 39, 115–133. doi: 10.1080/13682820310001616985
- Montgomery, J. W., and Leonard, L. B. (1998). Real-time inflectional processing by children with specific language impairment: effects of phonetic substance. *J. Speech Lang. Hear. Res.* 41, 1432–1443. doi: 10.1044/jslhr.41.06.1432
- Murphy, E., and Benitez-Burraco, A. (2016). Bridging the gap between genes and language deficits in schizophrenia: an oscillopathic approach. *Front. Hum. Neurosci.* 10:422. doi: 10.3389/fnhum.2016.00422
- O'Hara, M., and Johnston, J. (1997). Syntactic bootstrapping in children with specific language impairment. *Euro. J. Disord. Commun.* 32, 189–205.
- Owen Van Horne, A. J., and Lin, S. (2011). Cognitive state verbs and complement clauses in children with SLI and their typically developing peers. *Clin. Linguist. Phonet.* 25, 881–898. doi: 10.3109/02699206.2011.582226
- Penke, M. (2015). "Syntax and language disorders," in *Syntax — An International Hand book of Contemporary Research*, ed A. Alexiadou and T. Kiss (Berlin: Walter de Gruyter), 1833–1874.
- Perovic, A. (2001). Binding principles in Down syndrome. *UCL Work. Pap. Linguist.* 13, 423–455.
- Perovic, A., and Janke, V. (2013). Issues in the acquisition of binding, control and raising in high-functioning children with autism. *UCL Work. Pap. Linguist.* 25, 131–143.
- Perovic, A., Modyanova, N., and Wexler, K. (2013). Comparison of grammar in neurodevelopmental disorders: the case of binding in Williams syndrome and autism with and without language impairment. *Lang. Acquisit.* 20, 133–154. doi: 10.1080/10489223.2013.766742
- Petinou, K., and Terzi, A. (2002). Clitic misplacement among normally developing children and children with specific language impairment and the status of Infl heads. *Lang. Acquisit.* 10, 1–28. doi: 10.1207/S15327817L1001\_1
- Phillips, C. (2005). Electrophysiology in the study of developmental language impairments: prospects and challenges for a top-down approach. *Appl. Psycholinguist.* 26, 79–96. doi: 10.1017/S0142716405050071
- Reder, L. M., and Kusbit, G. W. (1991). Locus of the Moses illusion: imperfect encoding, retrieval, or match? *J. Mem. Lang.* 30, 385–406. doi: 10.1016/0749-596X(91)90013-A
- Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., et al. (2014). Specific language impairment: a convenient label for whom? *Int. J. Lang. Commun. Disord.* 49, 416–451. doi: 10.1111/1460-6984.12102
- Restrepo, M. A., and Gutierrez-Clellen, V. F. (2001). Article use in Spanish-speaking children with specific language impairment. *J. Child Lang.* 28, 433–452. doi: 10.1017/S0305000901004706
- Riches, N. G. (2012). Sentence repetition in children with specific language impairment: an investigation of underlying mechanisms. *Int. J. Lang. Commun. Disord.* 47, 499–510. doi: 10.1111/j.1460-6984.2012.00158.x
- Riches, N. G., Loucas, T., Baird, G., Charman, T., and Simonoff, E. (2010). Sentence repetition in adolescents with specific language impairments and autism: an investigation of complex syntax. *Int. J. Lang. Commun. Disord.* 45, 47–60. doi: 10.3109/13682820802647676
- Rondal, J.-A. (2007). "Language rehabilitation," in *Therapies and Rehabilitation in Down Syndrome*, eds J.-A. Rondal and A. Rasore Quartino (London: John Wiley and Sons), 63–89.
- Sanz-Torrent, M., Serrat, E., Andreu, L., and Serra, M. (2008). Verb morphology in Catalan and Spanish in children with specific language impairment: a developmental study. *Clin. Linguist. Phonet.* 22, 459–474. doi: 10.1080/02699200801892959
- Schuele, C. M., and Dykes, J. C. (2005). Complex syntax acquisition: a longitudinal case study of a child with specific language impairment. *Clin. Linguist. Phonet.* 19, 295–318. doi: 10.1080/02699200410001703709
- Shield, A., and Meier, R. P. (2012). Palm reversal errors in native-signing children with autism. *J. Commun. Disord.* 45, 439–454. doi: 10.1016/j.jcomdis.2012.08.004
- Shield, A., and Meier, R. P. (2014). Personal pronoun avoidance in deaf children with autism. *Proc. BUCLD* 38, 403–415.
- Spanoudis, G., Natsopoulos, D., and Panayiotou, G. (2007). Mental verbs and pragmatic language difficulties. *Int. J. Lang. Commun. Disord.* 42, 487–504. doi: 10.1080/13682820601010027
- Sprouse, J., and Almeida, D. (2013). "The role of experimental syntax in an integrated cognitive science of language," in *The Cambridge Hand book of Biolinguistics*, eds C. Boeckx and K. K. Grohmann (Cambridge: Cambridge University Press), 181–202.
- Stavrakaki, S., and van der Lely, H. (2010). Production and comprehension of pronouns by Greek children with specific language impairment. *Br. J. Dev. Psychol.* 28, 189–216. doi: 10.1348/026151010X485269
- Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile autism. *J. Autism Dev. Disord.* 11, 45–56. doi: 10.1007/BF01531340
- Terzi, A. (2005). "Comparative syntax and language disorders," in *The Oxford Hand Book of Comparative Syntax*, eds G. Cinque and R. Kayne (Oxford: Oxford University Press), 110–137.
- Terzi, A., Marinis, T., and Francis, K. (2016). The interface of syntax with pragmatics and prosody in children with autism spectrum disorders. *J. Autism Dev. Disord.* 46, 2692–2706. doi: 10.1007/s10803-016-2811-8
- Terzi, A., Marinis, T., Francis, K., and Kotsopoulou, A. (2012). Crosslinguistic differences in autistic children's comprehension of pronouns: English vs. Greek. *Proc. BUCLD* 36, 607–619.
- Terzi, A., Marinis, T., Kotsopoulou, A., and Francis, K. (2014). Grammatical abilities of Greek-speaking children with autism. *Lang. Acquisit.* 21, 4–44. doi: 10.1080/10489223.2013.855216
- Terzi, A., Zafeiri, A., Marinis, T., and Francis, K. (2017). Object clitics in the narratives of high-functioning children with autism. *Proc. BUCLD* 41, 637–650.
- Theodorou, E., and Grohmann, K. K. (2015). Object clitics in Cypriot Greek children with SLI. *Lingua* 161, 144–158. doi: 10.1016/j.lingua.2014.11.011
- Tomas, E., Demuth, K., Smith-Lock, K. M., and Petocz, P. (2015). Phonological and morphophonological effects on grammatical development in children with specific language impairment. *Int. J. Lang. Commun. Disord.* 50, 516–528. doi: 10.1111/1460-6984.12152
- Toro, R., Konyukh, M., Delorme, R., Leblond, C., Chaste, P., Fauchereau, F., et al. (2010). Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. *Trends Genet.* 26, 363–372. doi: 10.1016/j.tig.2010.05.007
- Tsakiridou, M. (2006). The linguistic profile of Down's syndrome subjects: evidence from wh-movement construction. *SOAS Work. Pap. Linguist.* 14, 227–248.
- Tsimpli, I. M., Kambanaros, M., and Grohmann, K. K. (2017a). "Language pathology," in *The Oxford Hand Book of Universal Grammar*, ed I. G. Roberts (Oxford: Oxford University Press), 486–508.
- Tsimpli, I. M., Peristeri, E., and Andreou, M. (2017b). Object clitic production in monolingual and bilingual children with Specific Language Impairment: a comparison between elicited production and narratives. *Linguist. Approac. Bilingual.* 7, 394–430. doi: 10.1075/lab.15025.tsi
- Tsimpli, I. M., and Stavrakaki, S. (1999). The effects of a morphosyntactic deficit in the determiner system: the case of a Greek SLI child. *Lingua* 108, 31–85. doi: 10.1016/S0024-3841(98)00041-2
- van den Heuvel, M. P., and Sporns, O. (2013). Network hubs in the human brain. *Trends Cogn. Sci.* 17, 683–696. doi: 10.1016/j.tics.2013.09.012

- van der Lely, H. K. J. (2005). Domain-specific cognitive systems: insight from Grammatical-SLI. *Trends Cogn. Sci.* 9, 53–59. doi: 10.1016/j.tics.2004.12.002
- Verhoeven, J. S., Rommel, N., Prodi, E., Leemans, A., Zink, I., Vandewalle, E., et al. (2012). Is there a common neuroanatomical substrate of language deficit between autism spectrum disorder and specific language impairment? *Cereb. Cortex* 22, 2263–2271. doi: 10.1093/cercor/bhr292
- Volden, J., Coolican, J., Garon, N., White, J., and Bryson, S. (2009). Pragmatic language in autism spectrum disorder: relationships to measures of ability and disability. *J. Autism Dev. Disord.* 39, 388–393. doi: 10.1007/s10803-008-0618-y
- Wittke, K., Mastergeorge, A. M., Ozonoff, S., Rogers, S. J., and Naigles, L. R. (2017). Grammatical language impairment in autism spectrum disorder:

exploring language phenotypes beyond standardized testing. *Front. Psychol.* 8:532. doi: 10.3389/fpsyg.2017.00532

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Leivada, Kambanaros and Grohmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Relationship between Syntactic Satiation and Syntactic Priming: A First Look

Monica L. Do\* and Elsi Kaiser

Department of Linguistics, University of Southern California, Los Angeles, CA, United States

## OPEN ACCESS

### Edited by:

Aritz Irurtzun,  
Centre National de la Recherche  
Scientifique (CNRS), France

### Reviewed by:

Grant Goodall,  
University of California, San Diego,  
United States  
Mikel Santesteban,  
University of the Basque Country,  
Spain

### \*Correspondence:

Monica L. Do  
monicado@usc.edu

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 17 May 2017

**Accepted:** 04 October 2017

**Published:** 25 October 2017

### Citation:

Do ML and Kaiser E (2017)  
The Relationship between Syntactic  
Satiation and Syntactic Priming:  
A First Look. *Front. Psychol.* 8:1851.  
doi: 10.3389/fpsyg.2017.01851

Syntactic satiation is the phenomenon where some sentences that initially seem ungrammatical appear more acceptable after repeated exposures (Snyder, 2000). We investigated satiation by manipulating two factors known to affect syntactic priming, a phenomenon where recent exposure to a grammatical structure facilitates subsequent processing of that structure (Bock, 1986). Specifically, we manipulated (i) *Proximity of exposure* (number of sentences between primes and targets) and (ii) *Lexical repetition* (type of phrase repeated across primes and targets). Experiment 1 investigated whether acceptability ratings of Complex-NP Constraint (CNPC) and Subject islands improve as consequence of these variables. If so, priming and satiation may be linked. When primes were separated from targets by one sentence, CNPC islands' acceptability was improved by a preceding island of the same type, but Subject islands' acceptability was not. When prime-target pairs were separated by five sentences, we found no improvement for either island type. Experiment 2 asked whether improvements in Experiment 1 reflected online processing or offline end-of-sentence effects. We used a self-paced reading paradigm to diagnose online structure-building and processing facilitation (Ivanova et al., 2012a) during processing. We found priming for Subject islands when primes and targets were close together, but not when they were further apart. No effects were detected when CNPC islands were close together, but there was a localized effect when sentences were further apart. The disjunction between Experiments 1 and 2 suggests repetition of the structure in Subject islands facilitated online processing but did not 'spill over' to acceptability ratings. Meanwhile, results for CNPC islands suggest that acceptability rating improvements in Experiment 1 may be driven by factors distinct from online processing facilitation. Together, our experiments show that satiation may not be a one-size-fit-all phenomenon but, instead, appears to manifest itself differently for different types of structures. Priming is possible and may be linked to satiation in *some* purportedly "unbuildable" structures (e.g., Subject islands), but not for all types (e.g., CNPC islands). Despite this, it appears that while the types of mechanisms targeting different island types are distinct, they are nevertheless similarly sensitive to the proximity between individual exposures.

**Keywords:** satiation, syntactic priming, island effects, processing difficulty, experimental syntax, acceptability judgments

## INTRODUCTION

Syntactic satiation is the phenomenon where some sentences that are “initially judged ungrammatical begin to sound increasingly acceptable” after repeated exposures (Snyder, 2000, p. 575). Anecdotally, this phenomenon is not new; most linguists have, at one time or another, fallen victim to “linguists’ disease.” Experimentally, though, evidence for satiation has yielded mixed results. So, while prior work has laid the groundwork for investigation, a number of fundamental questions remain – including the issue of which structures can/cannot satiate. Consequently, answering the subsequent questions of what mechanism and what factors underlie satiation has been challenging.

Existing work suggests that *only certain* syntactic violations satiate, while others are consistently perceived as unacceptable despite repeated exposure (e.g., Snyder, 2000; Sprouse, 2009). These structural asymmetries show that this poorly understood phenomenon has far-reaching implications for linguistic methodology (e.g., the design of acceptability judgment studies), for linguistic theories (e.g., the relative strength and status of syntactic violations, etc.), and for language processing theories (e.g., how the processor mentally represents ungrammatical sentences).

The present work investigates syntactic satiation from a new methodological and theoretical angle by manipulating variables known to affect a similar – though comparatively well-attested and better-understood phenomenon – known as syntactic priming (a.k.a. structural priming). Specifically, syntactic priming is a phenomenon where recent exposure to a given structure facilitates subsequent processing of that same structure (Bock, 1986; see Branigan, 2007 for a review). For instance, if a speaker has been recently exposed to a passive sentence (e.g., “The cat was chased by the dog”; the prime), she is more likely to produce another passive sentence (the target) the next time she is faced with a choice between an active and a passive structure (e.g., Bock, 1986).

The two phenomena of priming and satiation appear to resemble each other: In both cases, it’s *exposure* that influences how structures are processed. Despite this similarity, though, the literatures on priming and satiation have developed in relative isolation from one another. This may be partly due to differences in their methodological traditions. Priming, for instance, has been investigated almost exclusively with grammatical sentences (but see Kaschak and Glenberg, 2004; Ivanova et al., 2012a,b, 2017; etc.), often by means of production-oriented methods where the dependent variable is the proportion of trials on which a participant produces the primed structure. There have also been comprehension-oriented studies of priming (see Tooley and Traxler, 2010 for review), where the dependent variable is often ease of processing (as measured by eye-tracking, ERP, self-paced reading, etc.). Satiation, by contrast, has used offline acceptability judgments to see whether increased exposure improves the acceptability of ungrammatical sentences. Prior work on satiation has not made any direct claims about ease of processing for these ungrammatical sentences. Consequently, the broader relationship between priming and

satiation has been one of ‘apples and oranges’ as the potential relationship between these two phenomena has largely been overlooked.

Our work makes a first attempt at bridging these fields by using a priming-style design to investigate the mechanisms that may underlie satiation in two structures said to be ungrammatical in English, Complex Noun-Phrase Constraint (CNPC) islands and Subject islands. We present two experiments which approach satiation in a new way by manipulating two factors – namely (a) the proximity of prime and target sentences, and (b) the type of lexical repetition that occurs between them – known to affect syntactic priming.

Experiment 1 applies those factors to an offline acceptability rating task to test for rating improvements in CNPC and Subject islands. Acceptability ratings showed that CNPC islands were improved by a preceding CNPC structure. Subject islands, by contrast, did not appear to be affected by our manipulations. Moreover, improvements in CNPC islands occurred when primes and targets were separated by one intervening sentence, but not when sentences were separated by five interveners. Experiment 1 results suggest that priming may be linked to satiation, but that its effects may be dependent on the type of syntactic structure and the proximity of exposure between prime and target sentences.

Experiment 2 used word-by-word self-paced reading times to investigate whether acceptability rating improvements from Experiment 1 corresponded to processing facilitation during moment-by-moment comprehension. However, we first conducted a stop-being-grammatical-task, in order to (i) address potential concerns regarding the point at which readers perceive CNPC islands and Subject islands as being ungrammatical, and to (ii) guide the interpretation of the self-paced reading results in Experiment 2. In Experiment 2, in contrast to the offline acceptability ratings, *online* reading time measures detected priming in Subject islands: Reading times for Subject islands were faster when participants had just seen another Subject island, but only when primes and targets were close together. Surprisingly, despite offline rating improvements, we found no priming (no reading time facilitation) for CNPC islands in Experiment 2 when primes and targets were close together. We observed a priming effect localized to one word when CNPC islands were separated by five sentences.

Together, our results suggest that satiation may be a more nuanced phenomenon than previously thought: It appears to be dependent on the type of structure under investigation and its observability depends on the method used to investigate it. Consistent differences between CNPC and Subject islands in Experiments 1 and 2 lead us to believe that what has been viewed as a unified phenomenon of ‘satiation’ in both CNPC and Subject islands may not be unified after all: We may be dealing with two different phenomena that are only be superficially similar. Based on our results, we suggest that different mechanisms may be at work during the processing of CNPC and Subject islands. Our results also suggest that the proximity between individual exposures plays a role in both the offline acceptability and online comprehension of these island types.

## Syntactic Satiation

Work in syntactic satiation has typically focused on ‘island’ structures (ex. 3–4), *wh*-questions which are ungrammatical in English because they are said to violate constraints governing the movement of *wh*-phrases in English.

- (1) Who does Mary believe that John likes \_\_\_\_\_ ?
- (2) What does John know \_\_\_\_\_ fell on the floor?
- (3) \*Who does Mary believe [the claim] that John likes \_\_\_\_?  
[CNPC Island]
- (4) \*What does John know (that) [a bottle of \_\_\_\_] fell on the floor?  
[Subject Island]

More specifically, well-formed English questions (ex. 1–2) involve the creation of a ‘filler-gap dependency’ between the pronounced (the filler) and interpreted (the gap) *wh*-phrases. Though this dependency can span across multiple clauses, there are nevertheless conditions that govern the formation of the filler-gap dependency. When these conditions are violated, movement of the *wh*-filler to the front of the sentence is disallowed. In example (3), for instance, introducing a noun phrase (‘the claim’) between the filler and the gap embeds the *wh*-gap within a noun phrase from which *wh*-movement is not possible. Likewise, when the *wh*-gap appears within a subject phrase (‘a bottle of’), as in (4), the resulting sentence is ungrammatical. Because these phrases – namely, complex noun phrases and subjects, respectively – block the formation of *wh*-dependencies, they are considered ‘islands’ to extraction (here represented using brackets).

In the first experimental investigation of satiation, Snyder (2000) asked native English speakers to rate the grammaticality of several types of island structures.<sup>1</sup> Participants rated each sentence type a total of five times. To determine whether there had been any improvement in ratings, the number of ‘grammatical/acceptable’ responses in the first two vs. the last two exposures was compared. Sentences were said to improve, or ‘sate’, if there were more ‘grammatical/acceptable’ responses in the second half than in the first half of the study.

Notably, Snyder (2000) found that while some ungrammatical structures satiated, others did not.<sup>2</sup> However, more recent work has been unable to replicate some of these original findings. For instance, the satiation effects initially observed for CNPC islands have been replicated by some (e.g., Sag et al., 2007; Hofmeister and Sag, 2010; Goodall, 2011; Snyder, 2017 using acceptability ratings), but not by others (Hiramatsu, 2000 using Likert scale ratings; Sprouse, 2009 using magnitude estimation). In addition, related work by Sag et al. (2007) and Hofmeister and Sag (2010) investigated CNPC islands using self-paced reading where participants were asked to read two types of CNPC islands word-by-word: In the first type, *wh*-fillers were bare *wh*-phrases (e.g., ‘who’ or ‘what’), whereas in the second type, the *wh*-fillers were more informative *which*-NP phrases (e.g., ‘which convict’),

<sup>1</sup>Snyder (2000) tested seven different structures, finding satiation for *whether*-islands as well. But, because they do not allow us to incorporate repetition type as a factor, we exclude them from the current study. Snyder did not find satiation in want-for, that-trace, adjunct islands, or left branch sentences.

<sup>2</sup>Ross (1967) distinguishes between two sub-categories of CNPC violations: extraction out of a relative-clause NP and extraction out of a sentential complement NP. Following Snyder (2000) and others, we focus on only sentential complements.

which have been shown to be more acceptable (Karttunen, 1977; Maling and Zaenen, 1982; Pesetsky, 1987, 2000; etc.). Both Sag et al. (2007) and Hofmeister and Sag (2010) reported a similar result. Participants rated *which*-NP CNPC islands more acceptable than CNPC islands with bare *wh*-phrases. Additionally, reading times for CNPC islands with *which*-NPs did not differ from their grammatical, non-island counterparts. Results from both these studies were taken as evidence that under some circumstances, processing costs for CNPC islands could be drastically attenuated strictly by manipulating a single *processing*-related factor [(namely, the informativeness of the *wh*-element; but see Goodall (2015) for evidence of residual island effects even with highly informative filler phrases)]. We return to this point in the discussion.

Subject islands have been under similar debate. Although Snyder (2000) only showed a marginally significant effect of satiation, Hiramatsu (2000), Francom (2009), and Chaves and Dery (2014) have found significant satiation effects for Subject islands. Work by others, however, either replicated Snyder’s (2000) marginal effects (e.g., Snyder, 2017) or failed to detect satiation effects in these island types (e.g., Sprouse, 2009; Goodall, 2011; Crawford, 2012; etc.).

In sum, at issue is not only the question of (i) what mechanisms underlie satiation, but also the more fundamental question of (ii) whether what has been termed ‘satiation’ in CNPC and Subject islands is even the same phenomenon. In part because the basic facts of satiation remain unclear (e.g., there is no consensus regarding which structures do and do not satiate), it has been difficult to interpret what satiation as a phenomenon means both for experimental and for theoretical linguistics.

At a minimum, investigations into the phenomenon of satiation represent a methodological question for the design of acceptability judgment studies. For instance, a better understanding of the factors underlying satiation may have consequences for understanding individual variation in judgments, the number of times target items may be repeated, proximity of individual target items to each other, etc. Beyond that, satiation potentially implicates the interaction between grammatical constraints and how those constraints are mentally represented. This is particularly true in the case of grammatical violations, like CNPC and Subject islands, whose status in both the experimental and theoretical literature is still under debate.

## Syntactic Priming

Unlike satiation, syntactic priming – where exposure to a syntactic structure can facilitate subsequent processing of that same structure (Bock, 1986) – is a well-known and well-attested phenomenon. A large body of work (e.g., Bock, 1986; Branigan et al., 1995; Pickering and Branigan, 1998; Bock and Griffin, 2000) in priming has shown that speakers are better able to access structures (e.g., passive sentences) that they’ve previously been exposed to. And, though most of the research in priming focuses on production, similar priming effects have also been found in studies of comprehension. In general, the ability to facilitate access to recently exposed structures has been attributed to two complementary mechanisms that are not mutually exclusive (Hartsuiker et al., 2008): (1) residual activation of combinatorial

nodes in a syntactic structure (often lexically based), resulting in a short-lived priming effect (e.g., Pickering and Branigan, 1998; Branigan et al., 1999) and (2) Implicit learning of mappings between message-level representations and syntactic structures, resulting in a longer-term priming effect (Bock and Griffin, 2000; Chang et al., 2006; *inter alia*).

Residual activation accounts typically locate priming in the lexical units which connect to the larger syntactic structure (e.g., Pickering and Branigan, 1998; Branigan et al., 1999; Pickering et al., 2000; though see Scheepers, 2003). Since recent exposure momentarily increases the activation level of syntactic structures, priming occurs when the parser selects structures which are more active in memory, e.g., structures with higher residual activation levels. Because these accounts attribute priming to the moment-by-moment activation levels of particular lexicon-structure combinations, they also predict a short-term time course for priming (e.g., Roelofs, 1992; Pickering and Branigan, 1998). In particular, because the activation of lexical units is believed to decay quickly and automatically, priming effects are short-lived. Further, because residual activation accounts take priming to involve the links between lexical units and their larger syntactic structure, this account also predicts a stronger priming effect when prime and target sentences share lexical items (e.g., Pickering and Branigan, 1998; Cleland and Pickering, 2003). Indeed, this 'lexical boost' effect has been replicated in a number of production studies (e.g., Pickering and Branigan, 1998; Cleland and Pickering, 2003; Bernolet et al., 2013) and in nearly all comprehension studies (see Tooley and Traxler, 2010 for review).<sup>3</sup> But, other work has shown that priming can still occur absent lexical repetition in production (e.g., Pickering and Branigan, 1998; Scheepers, 2003; Kaschak and Glenberg, 2004; Hartsuiker et al., 2004) and comprehension (e.g., Luka and Barsalou, 2005; Thothathiri and Snedeker, 2008a,b; Traxler, 2008; Ivanova et al., 2012a,b).

A second mechanism contributing to structural priming – implicit learning – attributes priming to changes that occur independent of the lexicon; so, lexical repetition between prime and target sentences is not predicted to influence the strength of priming (Bock and Griffin, 2000; Chang et al., 2000, 2006; Bock et al., 2007). Rather, priming occurs as the result of cumulative, lasting learning from experience: Encountering a given message with a given structure reinforces learning of that meaning-to-message mapping. Consequently, the structure becomes more accessible the next time the processing system encounters the same type of message. Because priming under this account is the by-product of cumulative changes at the abstract structural level, priming is predicted to be relatively long-lasting (e.g., Hartsuiker and Kolk, 1998; Bock and Griffin, 2000; Bock et al., 2007; Hartsuiker et al., 2008). Work by Bock and Griffin (2000) measured the proportion of prepositional datives that participants produced after hearing a prepositional dative prime

(e.g., "A boy is giving an apple to a teacher.") or a double-object prime (e.g., "A boy is giving a teacher an apple."). To test the longevity of priming, they varied the number of unrelated sentences intervening between the prime and target structures. Consistent with prior work hinting at the persistence of priming, they found that effects could persist through as many as 10 intervening sentences.

The role of *ungrammatical* structures, though, is unclear. Most work in priming has focused on structural facilitation in the context of fully grammatical sentences – sentences whose structures *can* be mentally represented by the comprehender. Some researchers argue against the possibility of priming in ungrammatical sentences. For example, Sprouse (2007) suggests that priming "is predicated upon the existence of a licit representation. . . there should be no syntactic priming effect for ungrammatical structures" (Sprouse, 2007, p. 128). In contrast, other work (Kaschak and Glenberg, 2004; Luka and Barsalou, 2005; Ivanova et al., 2012a,b, 2017; etc.) has suggested that priming need not be limited to fully grammatical sentences.

At the lexical level, a series of experiments by Ivanova et al. (2012a,b, 2017) investigated if and how comprehenders build syntactic representations for anomalous ditransitive sentences (ex. 5a–b), when the verb is (a) a nonce word void of any semantic meaning, (b) a grammatically unacceptable verb, or (c) missing altogether. These anomalous sentences were compared against a fully grammatical counterpart (d).

- (5a) The waitress brunks the book to the monk/The waitress brunks the monk the book.
- (5b) The waitress exists the book to the monk/The waitress exists the monk the book.
- (5c) The waitress the book to the monk/The waitress the monk the book.
- (5d) The waitress gave the book to the monk/The waitress gave the monk the book.

Crucially, Ivanova et al. (2012a, 2017) used the presence/absence of syntactic priming effects (assessed via the proportion of participant-produced sentences matching the structure of the prime) to diagnose whether comprehenders had built syntactic representations for anomalous sentences.<sup>4</sup> They found evidence of structural priming – and thus the presence of abstract syntactic structure – with nonce-verb primes (5a), with illicit verb primes (5b) and even when the prime contained no verb (5c). Thus, work by Ivanova et al. (2012a, 2017) suggests that even when comprehenders encounter incomplete and/or ungrammatical sentences, they do not "abandon" the syntactic route altogether. In addition to using other available information, comprehenders do attempt to construct a representation for the sentence via syntax.

An open question, though, is whether findings from Ivanova et al. (2012a,b, 2017) can be straight-forwardly extended to

<sup>3</sup>While numerous production-based studies found priming even in the absence of lexical repetition, most comprehension-based studies found priming *only* when the prime and target have lexical overlap. An open question is whether this difference stems from the different tasks used to study priming in the two modalities or whether priming mechanisms in production and comprehension are fundamentally distinct (see Tooley and Traxler, 2010 for review).

<sup>4</sup>Ivanova et al. (2012a, 2017) also found that there was a priming 'boost' when verbs were the same. However, because priming was observed even when there was no lexical overlap, they concluded that even priming of anomalous sentences was lexically independent (but see Ivanova et al., 2012b).

account for structures as degraded as island structures (ex. 3–4). Anomalies in those works were largely localized to a single, albeit structurally important, lexical item – namely, the verb. Indeed, Ivanova et al. (2012b) themselves raise the question of whether their results may generalize to sentences where the locus of ungrammaticality extends beyond the level of individual lexical items – e.g., as in island structures (Ivanova et al., 2012b, p. 367).

Earlier work by Kaschak and Glenberg (2004) and Luka and Barsalou (2005) provide insights into what happens on the sentence level, although they did not test island structures. Specifically, Kaschak and Glenberg (2004) found priming-like effects in structures like ‘These vegetables need cooked,’ which are acceptable in some dialects, but ungrammatical in standard American English. In their experiment, half of the participants were exposed to the ‘needs’ structure during an initial training phase while the other half did not undergo training. Afterward, all participants were asked to read structurally similar sentences, such as ‘The valiant hero wants recognized for his courageous actions.’ Kaschak and Glenberg (2004) found faster word-by-word reading times for the novel ‘wants’ structures only for participants who had participated in the training session. This, they argued, provided evidence that participants were “learning to comprehend” the novel structure via a new meaning-to-message mapping (e.g., through implicit learning). Similar work by Luka and Barsalou (2005) investigated priming in a variety of moderately ungrammatical structures (e.g., ‘I miss having any time to do anything,’ ‘Who did you hire because he said would work hard?’). Participants first read sentences that were structurally similar to the target sentences, and after a 5-min break, rate the acceptability of the target sentences. Luka and Barsalou (2005) found acceptability improvements in as little as one prior exposure to a structurally similar sentence.

Taken together, these results indicate that priming may, indeed, be possible even with structures that *initially* seem unacceptable. Nevertheless, because work examining priming with ungrammatical sentences is relatively new, the limits of this priming effect are still unclear and the mechanisms and/or processes that underlie priming in ungrammatical sentences are not yet well-understood. Moreover, prior work has tended to either look at only one specific kind of anomaly, or has grouped together various types of ungrammatical sentences without comparing them systematically. Thus, it is not yet known how generalizable prior findings are, or whether different kinds of ungrammaticality may pattern differently with regard to the possibility of priming.

## The Current Study

The current work uses methods established in priming research to guide investigations into satiation, and in so doing, aims to shed light on broader issues related to the representation of ungrammatical sentences. Given the parallels between syntactic satiation and syntactic priming – namely, that both are linked to increased exposure – it may be possible for the underlying mechanism(s) responsible for satiation to be related to those in priming. The current work aims to contribute to our understanding of satiation and priming in three ways:

(1) Traditional approaches to satiation compared acceptability judgments over the course of an entire experiment, looking at cumulative effects on a ‘global’ level. By contrast, we test for improvements between prime and target pairs – ‘local,’ exposure-by-exposure comparisons – to see how single exposures to an ungrammatical prime can influence the acceptability of the subsequent target. Given that satiation effects have been notoriously difficult to replicate, even when studies have used similar materials, similar methods, and/or similar analyses (see *Syntactic Satiation*), looking at satiation through the lens of priming may provide independent evidence for how to interpret the facts of satiation.

(2) Whether structure-building is possible at all for ungrammatical, potentially ‘unrepresentable’ sentences like CNPC and Subject islands is an open question. Following Ivanova et al. (2012a,b, 2017), we use the presence of syntactic priming as a diagnostic for syntactic representation-building in cases where the input may be extremely degraded. In doing so, we examine not only the limits of representation-building, but also the ability of the processor to adapt to highly degraded input. Thus, our results also have implications for our understanding of the mental representations that underlie syntax, especially in the context of structures that may not be fully represented/representable in comprehenders’ minds.

(3) Finally, if comprehenders do, indeed, build syntactic representation of ungrammatical island sentences, an open question is to what extent processing of those representations may be similar to processing *grammatical* representations. We therefore “import” factors known to affect priming into our investigation of satiation to investigate the comparability of these two phenomena.

## EXPERIMENT 1: ACCEPTABILITY RATINGS

If proximity of exposure and lexical repetition – two factors known to modulate priming effects – can also increase the acceptability of CNPC and Subject islands, this might provide initial evidence that the same mechanisms underlying processing of grammatical sentences may play a role in how comprehenders’ evaluate notions of “(un)acceptability.” In other words, given that satiation is traditionally defined as increased acceptability, testing whether offline measures are influenced by processing-related factors is a key first step in determining whether priming and satiation are related.

Prior work in priming has shown that altering the number of sentences intervening between a prime and target can provide some insight into the mechanisms that contribute to priming. Because residual activation of a syntactic representation is short-lived, priming via this mechanism occurs when prime-target pairs are proximate, but not when they are further apart. By contrast, priming as an implicit learning effect appears to be long-lived (see *Syntactic Priming*). Thus, manipulating the *proximity* between prime and target sentences can shed light on one aspect of the underlying mechanism for satiation. We

operationalize this by changing the number of sentences (either one unrelated sentence, referred to as Lag1, or five unrelated sentences, referred to as Lag5) that intervene between a prime (the initial exposure sentence) and its target (the subsequent test sentence). Additionally, residual activation and implicit learning accounts with respect to the presence of a ‘lexical boost’ when primes and targets share lexical items critical to the syntactic structure (e.g., phrase heads, see *Syntactic Priming*). Therefore, we also manipulate *lexical repetition* between prime and target sentences by comparing repetition of a phrase crucial to the island-forming structure vs. repetition of lexical items unrelated to the island itself.

## Materials and Methods

### Participants

Eighty-four adult American English speakers, recruited via Amazon Mechanical Turk and paid \$2 (Lag1 group) or \$3 (Lag5 group), were included in the final analyses ( $n_{\text{Lag1}} = 40$ ,  $n_{\text{Lag5}} = 44^5$ ).

### Procedure

Participants saw sentences one at a time and rated how “natural or unnatural” each sentence “intuitively” sounded to them using a scale of 1 = “Completely Unacceptable” to 5 = “Completely Acceptable.” They were asked to rate sentences without reference to previously seen items and backtracking was disabled. The study was conducted using Qualtrics<sup>6</sup> (version 2015; Qualtrics, Provo, UT).

### Design

The number of sentences separating each prime from its subsequent target was varied between subjects: Prime-target pairs were separated either by one unrelated sentence (Lag1) or by five unrelated sentences (Lag5). Crucially, the total number of prime-target pairs was the same across both Lag1 and Lag5 versions; only the number of sentences intervening between primes and their targets varied. Specifically, participants rated three sets of prime-target pairs per condition (Table 1), for a total of six pairs in each sentence type and 12 prime-target pairs altogether. Additionally, participants rated 54 or 126 filler/intervener sentences in Lag1 and Lag5, respectively; these did not include island-related violations. Moreover, to address concerns that participants might “give up on” or adopt a response equalization strategy (Sprouse, 2009), participants rated a roughly equal number of ungrammatical and grammatical sentences over the course of entire study.

For each sentence type, targets were held constant but prime sentences were manipulated such that primes and targets either lexically repeated (i) the island-forming DP blocking *wh*-extraction or (ii) a phrase unrelated to the island (the matrix verb in CNPC islands and adjunct expressions in Subject islands). These four repetition conditions (Table 1) were varied within

subjects and rotated using a standard Latin Square design. (Note that repetition types are not compared to a no-repetition baseline).

Finally, in order to prevent the possibility that a ‘target’ could also function as a ‘prime’ for subsequent sentences, individual pairs of primes and targets were separated by at least 10 unrelated sentences. Comprehension questions were also interspersed throughout the experiment to further increase the distance between pairs of primes and targets (and to ensure people paid attention).

We now make several notes regarding the construction of our materials. First, complex-NP phrases can sometimes be reanalyzed as a single constituent (e.g., “make the claim” can be reanalyzed as “claim”). In cases of reanalysis, these ungrammatical sentences become fully grammatical because the *wh*-filler is no longer extracted from within a CNPC island (Cinque, 1990; Davies and Dubinsky, 2003; etc.). To minimize the possibility of reanalysis, we chose TP-complements to the VP that did not seem easily reducible to a single VP. Additionally, work by Phillips (2006) has shown that positing a gap inside of Subject islands (parasitic gaps) is not only possible inside island structures but can also “rescue” otherwise ungrammatical sentences. However, as noted by Phillips, parasitic gapping may be limited to infinitivals, so we test only finite clauses where “gap creation [is] not attempted” (Phillips, 2006, p. 813). Finally, given that prior work has shown satiation even with bare *wh*-phrases (Chaves and Dery, 2014), we use only bare *wh*-phrases to avoid additional processing confounds associated with more informative *wh*-fillers (Sag et al., 2007; Hofmeister and Sag, 2010).

### Predictions

If the same factors known to influence priming – namely, the proximity between individual (prime-to-target) exposures and the type of lexical overlap between structures – produce higher acceptability ratings for target sentences than for primes, this suggests that acceptability ratings may be sensitive to the same factors that affect processing. Such a finding would provide reason to suspect that priming and satiation can be linked to the same underlying mechanisms. Alternatively, if we observe no rating improvements between primes and targets, this would not rule out the possibility of a relationship between satiation and priming, but would make any such relationship indirect.

In priming, the proximity of exposure between prime-target pairs has been used to distinguish between effects arising from short-term residual activation decay and/or longer-term effects arising from implicit learning. We use this same logic to investigate whether rating improvements (satiation) may be short- or long-term. If acceptability ratings from prime to target sentences improve (i.e., satiate) when primes and targets are close together (Lag1; one intervening sentence), but show small improvements or no improvements when they are far apart (Lag5: five intervening sentences), this may point to satiation being a short-lived effect that decays over time. But, if both lags show comparable rating improvements, this could point to satiation as a long-term effect analogous to implicit structure-learning.

<sup>5</sup>We also excluded 11 participants who were either inattentive or did not understand the task. These participants responded to more than one comprehension questions incorrectly, rated grammatical fillers as ‘Completely Unacceptable,’ and/or were very slow in completing the experiment.

<sup>6</sup><http://www.qualtrics.com>



**TABLE 1** | Sample sentences (primes and targets) used in Experiments 1 and 2.

Sentence type	Repetition type	Trial type	Example sentences
CNPC	Island	Prime	Who did Richard dispute the <u>claim</u> that the paparazzi stalked?
		Target	Who did John deny the <u>claim</u> that the princess married?
	Unrelated	Prime	Who did Richard <b>deny</b> the allegation that the paparazzi stalked?
		Target	Who did John <b>deny</b> the claim that the princess married?
Subject	Island	Prime	What did <u>opponents</u> of hang a giant banner at the capitol?
		Target	What did <u>opponents</u> of start a violent riot outside the mall?
	Unrelated	Prime	What did fans of hang a giant banner <b>outside the mall</b> ?
		Target	What did opponents of start a violent riot <b>outside the mall</b> ?

Finally, lexical repetition often elicits a (short-lived) strengthening of the priming effect. According to residual activation accounts, this is because lexical repetition facilitates access to previously built syntactic structures. If acceptability is also sensitive to lexical repetition, we might find an analogous acceptability-rating 'boost' in Lag1 (primes and targets are close together) when prime-target pairs share lexical items. In particular, we may see stronger effects when the head of the syntactic island is repeated – given the significance of the head noun in the island structure – than when phrases unrelated to the island are repeated.

## Results

### Data Analysis

We measured changes in acceptability ratings (on a five-point scale) from prime to target sentences in CNPC and Subject islands. All statistical analyses were performed on *z*-scores computed from each participants' mean response to all experimental items. This helped control for differences in how individual participants would approach the five-point scale. However, analyses over raw ratings showed the same basic pattern of results. For ease of visual interpretation, graphs show raw ratings, not *z*-scores.

Statistical analyses were done in R (version 3.3.2; R Core Team, 2016) using linear mixed-effects regression models from the lme4 package (Bates et al., 2015). The Lag1 and Lag5 groups were compared independently. In all analyses, we included Sentence type (CNPC or Subject islands), Repetition Type (head of the island or an unrelated phrase), and Trial Type (prime vs. target sentence) as well as their interactions as fixed effects. We also incorporated by-subjects and by-items adjustments to the slopes and intercepts, which were reduced using model comparison.<sup>7</sup> Effects were judged to be significant if  $|t| \geq 2$ .

### Acceptability Ratings for Lag1

Mean ratings for sentences in the Lag1 group are shown in **Figure 1**. Overall, CNPC islands were rated significantly higher than Subject islands ( $b = 0.09$ ,  $SE = 0.03$ ,  $|t| = 2.82$ ). Moreover, ratings for CNPC target sentences were higher than for primes

regardless of repetition type. By contrast, ratings for prime and target Subject island sentences do not differ.

Statistically, there was a significant effect of trial type ( $\beta = 0.05$ ,  $SE = 0.02$ ,  $|t| = 2.3$ ), but this was modulated by a marginal sentence-by-trial interaction ( $\beta = 0.09$ ,  $SE = 0.05$ ,  $|t| = 1.81$ ). The presence of the interaction effect suggests that priming does not occur across the board: Target sentences were more acceptable than primes in CNPC islands ( $\beta = 0.1$ ,  $SE = 0.04$ ,  $|t| = 2.67$ ), but not Subject islands ( $\beta = 0.01$ ,  $SE = 0.03$ ,  $|t| = 0.40$ ).

There was no significant main effect of repetition type ( $\beta = -0.01$ ,  $SE = 0.02$ ,  $|t| = 0.41$ ) and no significant interactions ( $|t|'s < 0.36$ ) involving repetition type: Lexically repeating the head noun of the island itself vs. a phrase unrelated to the island did not affect ratings.

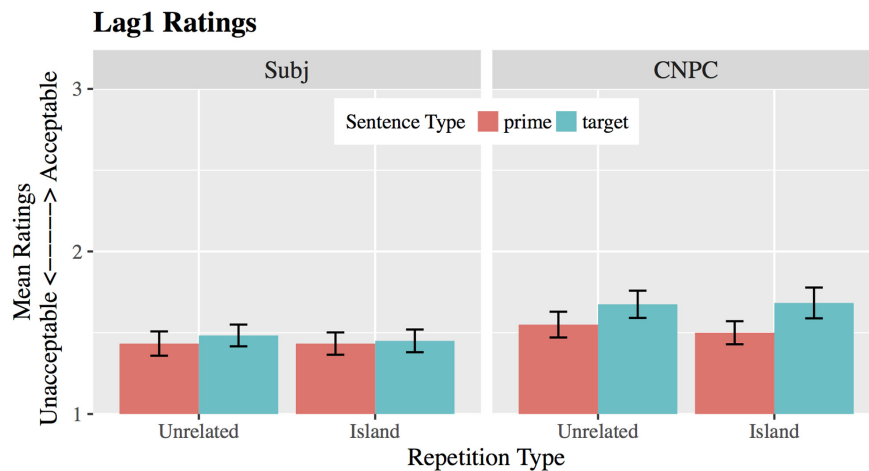
### Acceptability Ratings for Lag5

Ratings for prime and target sentences in Lag5 are shown in **Figure 2**. Mean ratings for CNPC islands were higher than for Subject islands, but this difference was only marginally reliable ( $\beta = 0.08$ ,  $SE = 0.04$ ,  $|t| = 1.91$ ). Unlike in Lag1, there was no significant effect of trial type ( $\beta = 0.03$ ,  $SE = 0.02$ ,  $|t| = 1.62$ ) and no significant sentence-by-trial interaction ( $\beta = 0.04$ ,  $SE = 0.05$ ,  $|t| = 0.91$ ): Ratings for target sentences did not significantly differ from prime sentences, either in CNPC or Subject islands. Lag5 also showed no main or interaction effects involving repetition type ( $|t|'s < 1.15$ ). Thus, in contrast to the improvements that we observed for CNPC islands in Lag1, no rating improvements were observed in Lag5, where primes and targets are separated by five intervening sentences.

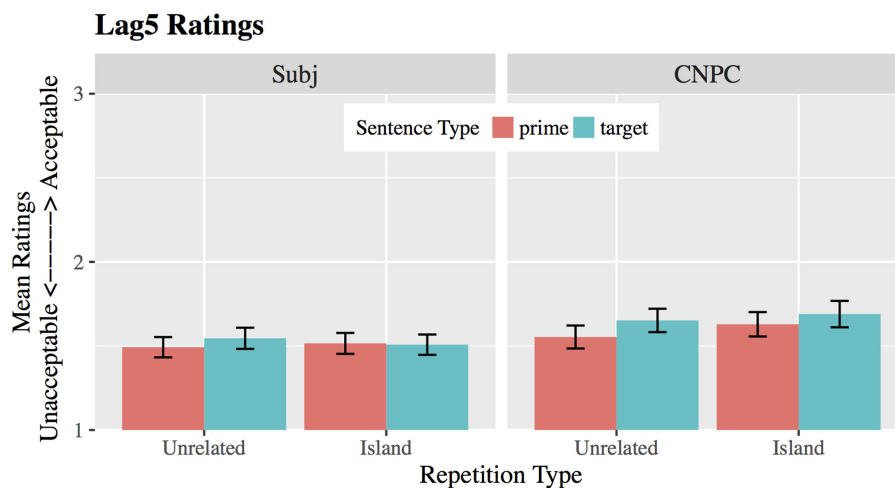
## Discussion

Experiment 1 investigated acceptability rating improvements for CNPC and Subject islands in prime-target pairs. While prior work in satiation has compared rating improvements over the course of an *entire* study, our priming-style (prime-target) design allowed us to test whether factors known to affect priming might also affect satiation similarly. If so, this might provide reason to suspect that priming and satiation share underlying mechanisms. We tested two factors: (1) lexical repetition and (2) proximity of exposure between the prime and target sentences. We varied lexical repetition such that primes and targets shared either the head of the island phrase or a phrase unrelated to the island. We predicted that repetition of the head of island phrases might produce a priming 'boost' akin to 'lexical boost' effects that have

<sup>7</sup>Random effects started out fully crossed and fully specified; they were reduced (starting with by-item effects) via model comparison, wherein only random effects that contributed significantly to the model ( $p > 0.05$ ) were included (Baayen et al., 2008).



**FIGURE 1** | Mean ratings for Complex-NP Constraint (CNPC) and Subject islands in Lag1. Raw scores are presented on a 5-point scale, where 1 = Completely Unacceptable and 5 = Completely Acceptable. Error bars represent  $\pm 1$  standard error. For visibility, we show only 1–3 points on the scale.



**FIGURE 2** | Mean ratings for CNPC and Subject islands in Lag5. Raw scores are presented on a 5-point scale, where 1 = Completely Unacceptable and 5 = Completely Acceptable. Error bars represent  $\pm 1$  standard error. For visibility, we show only 1–3 points on the scale.

been observed in priming work. In addition, we varied proximity of exposure by manipulating the number of unrelated sentences (one vs. five) between primes and targets, to probe whether potential acceptability improvements are short-term (e.g., from activation decay of structural representations) or long-term (e.g., as a result of implicit structural learning).

### Lexical Repetition

We found no effects involving the type of lexical items repeated across prime and target sentences. The finding that acceptability ratings show no lexical repetition effects might point to a fundamental difference in the mechanisms underlying satiation and priming. However, as previously mentioned in (see *Design*), we do not compare the types of lexical repetition to a baseline condition where primes and targets do not share any lexical items. Therefore, our results do not show that there is *no* effect

of lexical repetition – rather, our results provide evidence that the type of phrase that is lexically repeated does not affect the strength of priming for these sentence types. Furthermore, given that other work, including studies that examine priming in ungrammatical sentences (e.g., Kaschak and Glenberg, 2004; Luka and Barsalou, 2005; Ivanova et al., 2012a,b, 2017), found priming effects independent of ‘lexical boost’ effects, this should not be taken as evidence that priming is impossible either for CNPC or Subject islands.

### Overall Differences in Prime-to-Target Proximity

When primes and targets were separated by only one unrelated sentence (Lag1), participants rated CNPC targets as significantly more acceptable than their primes. But, when these same island types were separated by five sentences (Lag5), we found no effect of previous exposure. In other words, acceptability

ratings for CNPC islands satiated when sentences were close together, but not when they were further apart, suggesting that satiation is a short-lived effect that parallels what is predicted by lingering-activation accounts of syntactic priming (e.g., Pickering and Branigan, 1998; Branigan et al., 1999). Results from Experiment 1 therefore suggest that one factor that contributes to satiation may be a short-term priming effect that involves the lingering activation of structural representations which decay over time.<sup>8</sup>

### Overall Differences between CNPC and Subject Islands

We found that CNPC islands were generally more acceptable than Subject islands. More importantly, though, we also found that CNPC islands' acceptability ratings were *improved* by a proximate, preceding island (in Lag1), whereas Subject islands were not.

Our results provide initial evidence that satiation may be sensitive to the same factors known to affect priming. In other words, despite the indirect relationship between priming (a metric of processing ease) and acceptability ratings (a metric of well-formedness), there nevertheless appears to be a link between the two. However, our results also suggest that factors that affect priming do not seem to affect ratings across the board: They are in some way modulated by syntactic structure (e.g., CNPC island vs. Subject island). While CNPC islands were judged more acceptable in the context of a previously seen CNPC island, Subject islands did not benefit from a preceding Subject island.

### Differences between CNPC and Subject Islands: The Stop-Being-Grammatical Task

The results of Experiment 1 suggest that rating improvements (satiation) in CNPC islands are affected by the same factors that affect priming whereas ratings for Subject islands are not. However, so far we have focused on end-of-sentence acceptability ratings, which may not reflect the processes that occur as comprehenders incrementally process CNPC and Subject islands. To gain insights into the online, incremental processing of these two islands types, we used the self-paced reading paradigm in Experiment 2. But before turning to the reading-time data, we need to address a difference between CNPC islands and Subject islands that can have implications for our interpretation of the data – namely, the relative distance between the *wh*-gap and the head of the island phrase in CNPC vs. Subject islands. Specifically, in CNPC islands (ex. 3, repeated here as 6a),

the parser encounters the island-producing phrase ('the claim') earlier than the *wh*-gap (marked with \_\_\_\_\_) at the end of the clause. In contrast, in Subject islands (ex. 4, repeated here as 6b), the island phrase ('a bottle of \_\_\_\_\_') and the *wh*-gap (marked with \_\_\_\_\_) are fundamentally one and the same.

- (6a) CNPC Island: Who does Mary believe [the claim] that John likes \_\_\_\_\_?  
 (6b) Subject Island: What does John know (that) [a bottle of \_\_\_\_\_] fell on the floor?

If it is the presence of the *gap* site – not the island-producing phrase itself – that signals “ungrammaticality”, then comprehenders may treat CNPC islands as fully grammatical until they reach the sentence-final *wh*-gap. In other words, it could be that rating improvements observed for CNPC islands – and absent for Subject islands – may not be attributable to any theoretical differences between the two islands, but simply to the fact that CNPC islands effectively *appear* grammatical for a longer amount of time.

To test this possibility, we investigate the earliest point at which comprehenders perceive CNPC islands to be ungrammatical. At the same time, this 'stop-being-grammatical' task also contributes to our broader goal of probing the relationship between what has been a predominantly off-line phenomena (satiation) and online facilitation effects, by providing new information about acceptability judgments at different points over the course of the sentence.

Twenty-seven native American English speakers were recruited via Amazon Mechanical Turk to participate in the stop-being-grammatical task, modeled after the stop-making-sense task (Boland et al., 1990, 1995; etc.) in Qualtrics<sup>9</sup> (version 2017; Qualtrics, Provo, UT).

Two CNPC and two Subject islands and six filler sentences were randomly selected from Experiment 1. (Note that while Subject islands are included, they are not of interest because of the island and *wh*-gap essentially occur simultaneously. They are shown for comparison in **Figure 3**, but statistics are reported only for CNPC islands). Sentences were presented to participants in successive fragments, such that each new fragment added one more word to the end of the sentence. The initial fragment consisted of the first two words (e.g., 'Who did,' or 'What did') and subsequent fragments increased by one word. So, if participants initially saw "Who did Brandon," the next fragment would be "Who did Brandon make"; the fragment after would contain one more word until the last word of the sentence was reached. Participants had 45 s to determine ('Yes'/'No') whether each fragment could be continued to make an "acceptable"/"possible" sentence of English.

**Figure 3** shows the cumulative percentage of 'No' responses at each word position.<sup>10</sup> At word 5 (determiner 'the' in CNPC islands, matrix verb in Subject islands), the number of 'No'

<sup>8</sup>Even though we discuss numerical differences between Lag1 and Lag5, between-group effects were not compared directly. Because our study is the first of its kind to explore the links between satiation and priming in this way, while also comparing different island types, it was not designed with the statistical power to detect a 3-way, between-subjects interaction. Between-subjects effects are difficult to detect, especially without sufficient statistical power. The situation is further complicated by the well-known observation that structural priming effects are relatively small, and the fact that the effect of interest is a three-way interaction between sentence type, trial type and lag. Even though our sample size is in line with current psycholinguistic work, we do not expect to be able to detect this kind of interaction. Additional exploratory analyses suggest that doing so would require a sample size much larger than what is standard in psycholinguistics. Nevertheless, we acknowledge that further work requires a more vigorous focus on effect sizes, power, and sufficient sample size.

<sup>9</sup><http://www.qualtrics.com>

<sup>10</sup>The cumulative percentage of responses at any position is fundamentally dependent on the number of 'No' responses prior to that position. To minimize this dependence, we also calculate adjusted percentages such that the number of 'No' responses was out of the total of "remaining possible no" at each position (Boland et al., 1990). Adjusted percentages showed the same pattern.

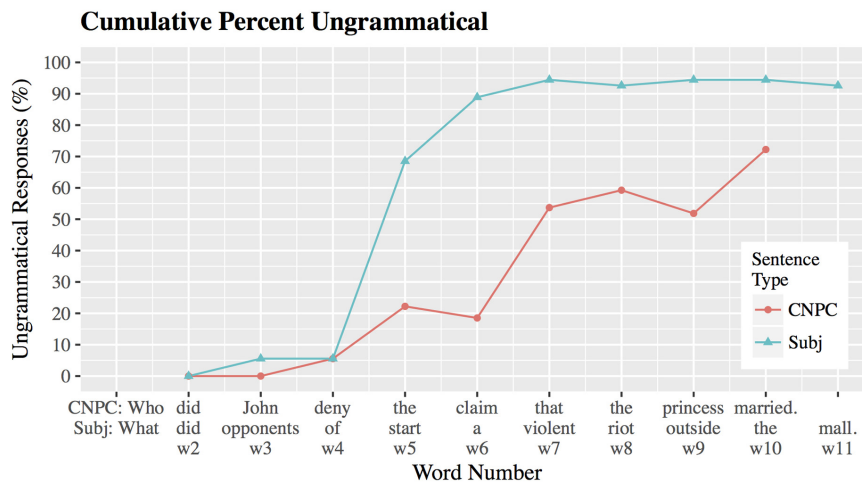


Figure 3: Cumulative percentage of 'No' (fragment is ungrammatical) at each word in the stop-being-grammatical task.

FIGURE 3 | Cumulative percentage of 'No' (fragment is ungrammatical) at each word in the stop-being-grammatical task.

responses increases for both sentence types; but at different rates for Subject vs. CNPC islands. Notably, at word 5, 70% of participants judge Subject islands to be ungrammatical with 90% of participants concurring by word 6. By contrast, although some participants judge CNPC islands to be ungrammatical at word 5, the majority do not until word 7 (complementizer 'that'). Responses were analyzed using logistic mixed-effects regressions with random intercepts for subjects and items. We first compared responses word 4 (low rates of unacceptability) against responses at words 5 and 6 (increasing rates of unacceptability). We found a significant effect of word position for both CNPC ( $\beta = -1.88$ ,  $SE = 0.71$ ,  $|z| = 2.65$ ) and Subject islands ( $\beta = -4.56$ ,  $SE = 0.93$ ,  $|z| = -4.92$ ), meaning that the proportion of 'No' responses (i.e., ungrammatical responses) at word 4 was significantly lower than at words 5 and 6 for both island types. Contrasting words 5 and 6 yielded no significant differences for CNPC islands ( $\beta = 0.45$ ,  $SE = 0.68$ ,  $|z| = 0.67$ ), but we did find a significant increase from word 5 to word 6 in Subject islands ( $\beta = -2.18$ ,  $SE = 0.78$ ,  $|z| = -2.79$ ).<sup>11</sup>

Results from the stop-being-grammatical task suggest that judgments of (un)acceptability, like sentence processing itself, may proceed incrementally and 'unacceptability' is expected to begin around word 5 for both Subject and CNPC islands. More importantly, even if CNPC islands are arguably fully grammatical until the sentence-final *wh*-gap, comprehenders begin *perceiving* CNPC islands to be ungrammatical much earlier (around word 5, with a majority of comprehenders concurring by word 7). These findings argue against the potential concern that the CNPC-Subject island asymmetry in Experiment 1 was due to CNPC islands being perceived as grammatical/acceptable until

the gap site at the end of the sentence. Our results suggest that comprehenders do *not* wait for the *wh*-gap to 'decide' whether a sentence is ungrammatical.

## EXPERIMENT 2: SELF-PACED READING

Experiment 1 provided initial evidence that acceptability ratings might be tuned to the same factors that have been found to affect online processing. However, given that prior work on satiation has mainly used acceptability ratings it is not yet known whether (i) it is end-of-sentence, meta-linguistic reflection that causes rating improvements to 'kick in' or whether (ii) rating improvements reflect incremental, processing facilitation. For instance, in contexts as structurally degraded as island sentences, comprehenders may rely primarily on processes outside of syntactic structure-building (e.g., plausibility, discourse context, word order, etc.). If so, rating improvements may not correspond to the type of facilitation characteristic of structural priming. Alternatively, in line with what has been observed in structure-building for anomalous sentences (Ivanova et al., 2012a,b, 2017), comprehenders may nevertheless engage structural (re)integration processes even despite the type of ungrammaticality presented by island structures.

Therefore, Experiment 2 builds on Experiment 1 and the stop-being-grammatical task by directly testing whether the *online* processing of CNPC and Subject islands can be facilitated by a prior exposure. We use the self-paced reading paradigm to probe reading time slowdowns, which often stem from processing difficulty. In doing so, we probe the source of the rating improvements observed in Experiment 1, and by extension, determine whether offline rating improvements (i.e., satiation) correspond to online processing facilitation effects (i.e., priming). If recent exposure to ungrammatical structures can decrease the processing costs associated with ungrammatical structures, we

<sup>11</sup>While a statistically significant difference between words 5 and 6 in Subject islands is interesting, it is ultimately irrelevant to the central aims of the stop-being-grammatical task. Namely, to determine the *first* point at which the sentence becomes ungrammatical. Therefore, we do not discuss reasons for difference between words 5 and 6 in Subject islands.

might expect faster reading times for target sentences relative to their prime counterparts, which would not have the benefit of a recent facilitating exposure.

## Predictions

### Lexical Repetition

Experiment 1 showed no effect of lexical repetition, so we do not expect differences here. We collapse repetition types in Experiment 2 to increase statistical power.<sup>12</sup>

### Proximity of Exposure

Experiment 1 found that for CNPC islands, acceptability ratings improved when primes and targets were proximate (Lag1) but not when they were further apart (Lag5). This suggests that satiation may be a by-product of short-term lingering activation. If these short-term effects can be linked to those observed in short-term priming, we expect reading times to improve from primes to targets when sentences are close together (Lag1), but not when they are further apart (Lag5). But, it may also be possible that while rating improvements (satiation) are short-term, online facilitation in island sentences is the result of a more long-term priming mechanism, such as implicit learning. In the latter case, we expect prime-to-target reading time improvements regardless of whether prime and target sentences are separated by one or by five intervening sentences (Lag1 and Lag5).

### Sentence Types

Based on prior work (Kluender and Kutas, 1993; Phillips, 2006; Sag et al., 2007; Hofmeister and Sag, 2010; etc.), we expect processing difficulty (gauged via reading time slowdowns) to arise at word 5 for CNPC and Subject islands (see **Table 2**), but crucially, for different reasons. In both cases, the parser begins actively searching for a *wh*-gap as soon as it encounters the sentence-initial *wh*-phrase ('Who'/'What'; Crain and Fodor, 1985; Frazier and Clifton, 1989; Gibson and Hickok, 1993; etc.). In CNPC islands, the processing difficulty expected at word 5 can be attributed to what is known as the filled-gap effect: The parser posits a gap for the *wh*-filler at the first possible position, word 5 (**Table 2**); but, when it encounters the head of the island phrase ('the') here, the parser realizes that this is not a possible position for the *wh*-gap and must revise its initial parse. We also expect a secondary site of processing difficulty at word 7, where the parser encounters the complementizer (e.g., 'that'). Here, because the complementizer signals the end of the previous clause and because there was no available gap position in the initial clause, the parser should recognize that the *wh*-filler has been extracted from within an embedded clause headed by a complex-NP – in other words, that the *wh*-filler has been extracted from within a CNPC island. Thus, the expected processing difficulty at word 7 would correspond to the point where the parser has recognized the illicit, ungrammatical extraction. Indeed, these predictions are in line with what we observe in the stop-being-grammatical

task (see *Differences between CNPC and Subject Islands: The Stop Being Grammatical Task*): Some comprehenders *begin* perceiving CNPC islands as unacceptable at word 5 with the majority of comprehenders judging CNPC islands to be unacceptable by word 7.

In the case of Subject islands, we also expect processing difficulty to begin at word 5. However, because the parser does not postulate gaps within finite islands (Phillips, 2006), any potential processing difficulties observed here cannot be due to the filled-gap effect. In Subject islands, word 5 is the point where the parser begins to recognize the ungrammatical extraction: When the parser encounters the preposition ('of') at word 4, it expects that another noun phrase will follow. When it instead encounters a verb ('start'), the parser realizes that the *wh*-filler has been extracted from within a subject phrase (i.e., a Subject island). Again, this is in line with where the majority of comprehenders in the stop-being-grammatical task (see *Differences between CNPC and Subject Islands: The Stop Being Grammatical Task*) judge Subject islands to be unacceptable.

Experiment 1 found lower ratings for Subject than for CNPC islands. Given this, one might be tempted to also predict that Subject islands might be read slower than CNPC islands. But, due to overall differences between the two sentences (e.g., word length, word frequency, etc.), we cannot compare the two sentence types directly. Rather, our comparison of interest is a sentence-by-trial *interaction*, measuring priming in CNPC vs. Subject islands that would signal this asymmetry in processing. In other words, finding that Subject and CNPC islands have different reading times (a main effect of sentence type) cannot help us to determine whether satiation and priming are linked to the same mechanisms. What is relevant is whether the same pattern of asymmetrical improvements between CNPC vs. Subject islands that was observed in Experiment 1 will also be present using an online metric. Only a sentence-by-trial interaction can speak to this asymmetry.

## Materials and Methods

### Participants

Thirty-four ( $n_{Lag1} = 18$ ;  $n_{Lag5} = 16$ ) native speakers of American English from the University of Southern California participated in Experiment 2. Participants received course credit or \$10 for participation. The experiment lasted roughly 45 min.

### Procedure

The study was conducted in a quiet room at the University of Southern California. Sentences were presented using Linger (D. Rohde, MIT; Rohde, 2010).

Participants were told that sentences would start out completely masked by dashes. They were instructed to read the sentences as quickly and carefully as possible, using the 'space bar' to reveal each word in the sentence one-by-one. After reading the last word in the sentence, participants saw a scale ranging from 1 (Completely Unacceptable) to 7 (Completely Acceptable), where they used the mouse to rate how each sentence "intuitively" sounded to them. Participants would intermittently see a comprehension question about the sentence they just read.

<sup>12</sup>At the request of a reviewer, additional analyses were carried out with repetition type as a fully crossed fixed effect. It was also included in the by-subject and by-item random effect structures. Model reduction was done as previously described. In both Lag1 and Lag5, there were no significant main or interaction effects involving repetition type during any portion of our region of interest.

**TABLE 2** | Sample Complex-NP Constraint (CNPC) and Subject island sentences with corresponding word numbers. Shaded region denotes region of interest.

Word number	w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	w11
CNPC:	Who	did	John	deny	the	claim	that	the	princess	married?	
Subj:	What	did	opponents	of	start	a	violent	riot	outside	the	mall?

## Design

Experiment 2 used the same materials as Experiment 1. Again, two versions of the study (Lag1 vs. Lag5) were tested between-subjects. To ensure that participants were paying attention, we asked them to provide acceptability ratings for every sentence presented. However, given the extreme task differences in Experiment 1 vs. Experiment 2, we did not expect results from this rating task to be meaningful or comparable (Sag et al., 2007; Hofmeister and Sag, 2010; Hofmeister et al., 2012a,b; etc.).<sup>13</sup> We report acceptability ratings for the sake of completeness, but they are not discussed further.

## Results

### Data Analysis

Reading times below 100 ms, above 3000 ms, and more than three standard deviations above the positional mean for each condition were excluded, affecting 2 and 1.7% of the data in Lag1 and Lag5, respectively. Our region of interest began at word 5 (Table 2) and extended for three additional words. Because the structure of CNPC and Subject islands do not parallel each other, we do not compare them directly. Consequently, our comparison of interest is a *sentence-by-trial interaction* that compares the degree to which reading times are facilitated across island types.

Results from Lag1 and Lag5 were analyzed independently using linear mixed-effects models (Bates et al., 2015) in R (R Core Team, 2016). We included sentence type, trial type, and their interaction as fixed effects predictors. Random effect structure was determined as in Experiment 1.

### Results from Lag1

Figure 4 and Table 3 show reading times for prime and target sentences in CNPC and Subject islands in Lag1. Except at w7, we find a significant main effect of sentence type ( $w5: \beta = 57.04, SE = 22.47, |t| = 2.54$ ;  $w6: \beta = 64.26, SE = 21.48, |t| = 2.99$ ;  $w7: \beta = 27.64, SE = 18.31, |t| = 1.51$ ;  $w8: \beta = 47.01, SE = 13.91, |t| = 3.38$ ), meaning that both primes and targets for Subject islands are read slower than CPNC islands. While expected, this effect is not informative given that differences between these islands can range from individual lexical items to broader structural differences.

At word 5, CNPC islands do not show any reading time slowdowns, even though results from our stop-being-grammatical task predicted a reading time increase at this point in the sentence. Reading times for Subject islands increase at w5, consistent with results from the stop-being-grammatical task. However, we do not detect a significant main effect of trial type

( $\beta = 22.80, SE = 20.69, |t| = 1.10$ ) or a significant sentence-by-trial interaction ( $\beta = -15.51, SE = 24.12, |t| = 0.64$ ), meaning that reading times for primes and targets did not differ for either sentence type.

At words 6 and 7, reading times for Subject islands improve as a result of recent exposure (i.e., priming) in Subject islands ( $w6: \Delta = 75.08$  ms;  $w7: \Delta = 61.50$ ms) but not for CNPC islands ( $w6: \Delta = 9.49$  ms;  $w7: \Delta = 8.30$  ms). This asymmetry in priming is corroborated by a significant sentence-by-trial interaction ( $w6: \beta = 59.16, SE = 30.40, |t| = 1.95$ ;  $w7: \beta = 50.89, SE = 25.79, |t| = 1.97$ ). Thus, seeing an initial Subject island facilitated processing of the subsequent Subject island. In CNPC islands, reading times for primes and targets did not differ from each other regardless of whether or not comprehenders had seen a preceding prime. Interestingly, even though a majority of participants in the stop-being-grammatical task (see *Differences between CNPC and Subject Islands: The Stop Being Grammatical Task*) rated CNPC islands as “ungrammatical” by word 7, we also find no reading time slowdown here.

After w8, reading times for Subject islands converge and appear indistinguishable. At w11, reading times increase, presumably as a result of sentence-final wrap-up effects. In CNPC islands, sentence-final wrap-up effects emerge at w10.

Recall that participants in the self-paced reading study were also asked to rate the acceptability of the sentences on a 7-point scale, to ensure they were playing attention. However, given the extreme task differences in Experiment 1 vs. 2, we did not expect these results to be meaningful (see *Design*). Analyses were conducted on z-scored acceptability ratings, but for ease of interpretability, we discuss raw scores.<sup>14</sup> Mean ratings for CNPC island primes and targets were 2.01 and 2.11, respectively; ratings for Subject island primes and targets were 1.88 and 1.75, respectively. As expected, there were no differences between CNPC and Subject islands overall ( $\beta = -0.08, SE = 0.08, |t| = 1.06$ ), no differences between primes vs. targets ( $\beta = 0.02, SE = 0.04, |t| = 0.46$ ), and no sentence-by-trial interaction ( $\beta = 0.09, SE = 0.08, |t| = 1.19$ ).

### Results from Lag5

Figure 5 and Table 4 show results for reading times in Lag5. At word 5, reading times for both Subject and CNPC islands increase, but they do not differ from each other ( $\beta = 19.35, SE = 16.73, |t| = 1.16$ ). (Recall that a main sentence type effect would not be interpretable in any case). However, we did find a significant main effect of trial type at word 5 ( $\beta = 32.65, SE = 16.68, |t| = 1.96$ ), meaning that Subject and CNPC prime sentences were read significantly slower than their target counterparts. There was no sentence-by-trial interaction at word

<sup>13</sup>Participants in Experiment 1 were shown the full sentences during the rating task. Participants in Experiment 2, however, first read the sentence word-by-word and then rated the sentence from memory after the sentence had disappeared from the screen.

<sup>14</sup>In all cases, analyses performed over raw score ratings showed the same pattern as z-scored ratings.

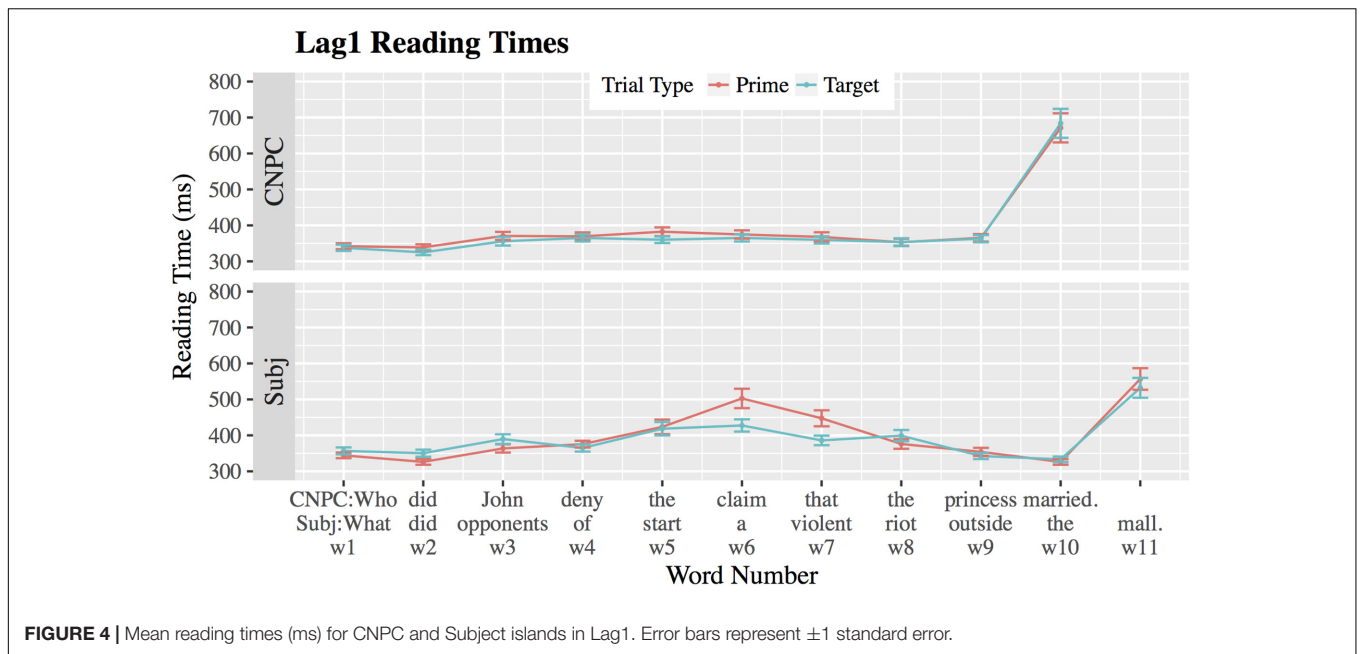


FIGURE 4 | Mean reading times (ms) for CNPC and Subject islands in Lag1. Error bars represent ±1 standard error.

TABLE 3 | Lag1 mean reading times for words in the region of interest.

Lag1: Reading times (ms)				
	Word 5	Word 6	Word 7	Word 8
CNPC_prime	382.44	374.73	368.03	353.16
CNPC_target	360.18	365.24	359.73	353.35
Subj_prime	423.68	502.69	447.52	376.10
Subj_target	418.54	427.61	386.02	399.23
Sig. effects detected	*Sentence Type	*Sentence Type *Sentence × Trial	*Sentence × Trial	*Sentence Type

Times are shown in milliseconds.

5 ( $\beta = -21.44$ ,  $SE = 23.69$ ,  $|t| = 0.91$ ), meaning that reading time differences between primes and targets were of the same magnitude regardless of sentence type.<sup>15</sup>

For all other words in the region of interest (w6–w8), we find only a significant effect of sentence type (w6:  $\beta = 81.33$ ,  $SE = 26.39$ ,  $|t| = 3.08$ ; w7:  $\beta = 56.21$ ,  $SE = 18.61$ ,  $|t| = 3.02$ ; w8:  $\beta = 51.63$ ,  $SE = 16.77$ ,  $|t| = 3.08$ ), meaning that Subject islands were read slower than CNPC islands. However, as previously noted, this comparison is not central to the aims of Experiment 2. We also find no main effect of trial type ( $|t|$ 's < 1.04), meaning that the difference in prime and target reading times observed at word 5 disappeared quickly. Crucially, the sentence-by-trial interaction previously observed in Lag1 was no longer detected from w6–w8. (Despite apparent graphical differences at word 6,

the sentence-by-trial interaction is not significant; it approaches marginal significance:  $\beta = 60.14$ ,  $SE = 37.26$ ,  $|t| = 1.614$ . For all other words,  $|t|$ 's < 1.44). At w10 and w11, reading times rise, presumably as a result of sentence-final wrap-up effects.

When we look at acceptability ratings in Lag5, we find that CNPC island ratings for primes and targets averaged 2.32 and 2.13, respectively, while Subject island ratings for primes and targets averaged 1.96 and 1.84, respectively. Unsurprisingly, CNPC and Subject islands did not differ from each other ( $\beta = -0.13$ ,  $SE = 0.09$ ,  $|t| = 1.44$ ); nor did primes and targets ( $\beta = 0.07$ ,  $SE = 0.05$ ,  $|t| = 1.44$ ). There was no sentence-by-trial interaction ( $\beta = -0.03$ ,  $SE = 0.09$ ,  $|t| = 0.29$ ).

## Discussion

Experiment 2 used an online measure – self-paced reading times – to investigate whether the acceptability rating improvements in Experiment 1 were related to on-line island processing effects. We tested for the presence of reading time improvements, indicative of processing facilitation, for CNPC and Subject islands when primes and targets were close together (Lag1) and when they were further apart (Lag5). Based on

<sup>15</sup>At the request of an reviewer, we performed separate analyses for CNPC and Subject islands at w5 with the hunch that prior analyses were insufficiently powered to detect the interaction effect. This subsequent analysis showed that main effect of trial type was primarily driven by CNPC islands. However, this does not impact the main claims of this paper. We provide additional discussion of the localized one-word priming effect for CNPC islands in Section “Discussion.” The lack of an effect for Subject islands at w5 is consistent with our claims that priming does not occur when primes and targets are further apart.

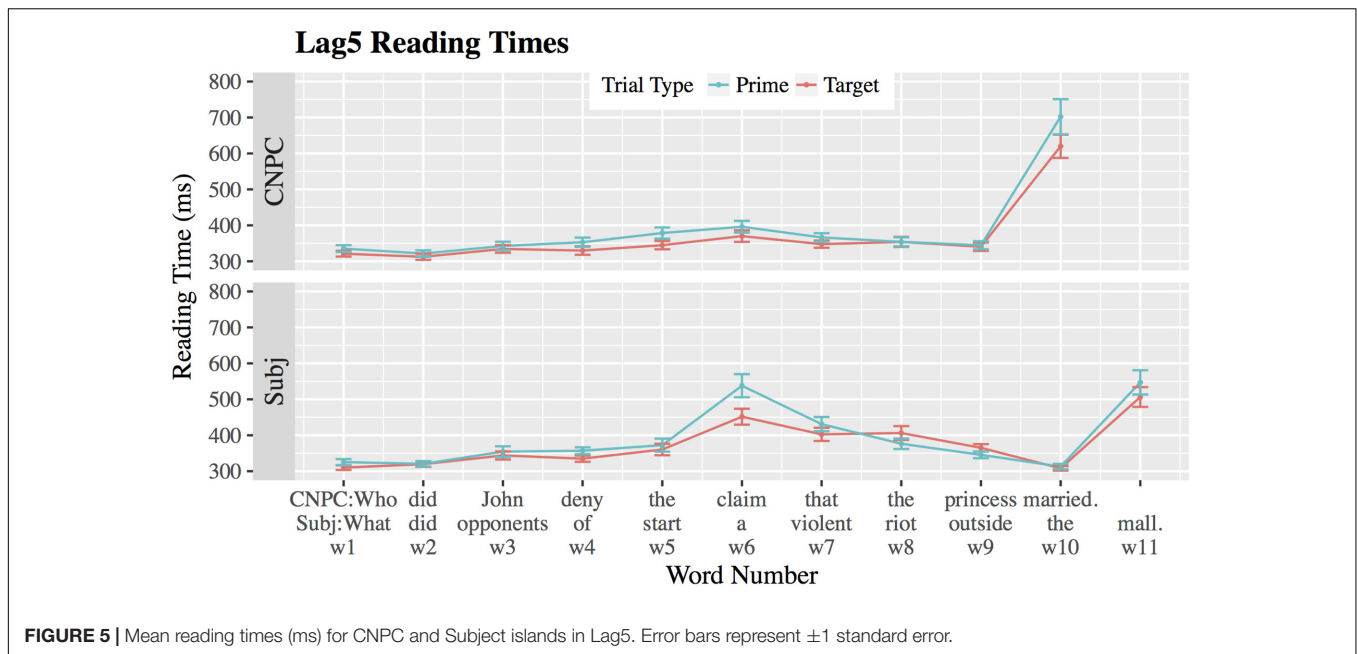


FIGURE 5 | Mean reading times (ms) for CNPC and Subject islands in Lag5. Error bars represent ±1 standard error.

TABLE 4 | Lag5 mean reading times for words in the region of interest.

Lag5: Reading times (ms)				
	Word 5	Word 6	Word 7	Word 8
CNPC_prime	378.66	396.12	366.34	354.13
CNPC_target	344.92	370.02	347.71	353.96
Subj_prime	372.49	537.87	431.00	376.46
Subj_target	360.43	451.67	402.53	406.34
Sig. effects detected	*Trial Type	*Sentence Type	*Sentence Type	*Sentence Type

Times are shown in milliseconds.

results from Experiment 1, we predicted that if the acceptability rating improvements found in CNPC islands (but not Subject islands) reflected online processing facilitation, we should find corresponding prime-to-target reading time facilitation in CNPC islands (but not Subject islands) in Experiment 2. We also investigate whether online facilitation effects in Experiment 2 were short- or long-term. If target sentences are read faster than prime sentences in Lag1, but not in Lag5, this would point toward a short-lived priming effect. But if reading times for targets in both Lag1 and Lag5 are faster than their primes, this would suggest a long-lasting effect.

Unlike in Experiment 1, which found no rating improvements for Subject islands regardless of proximity between prime and target sentences, Experiment 2 found faster reading times for target sentences when Subject islands were separated by only one intervening sentence (Lag1). This effect lasted through several words in our region of interest. When sentences were further apart (Lag5), we found a prime-to-target facilitation localized to only one word in the region of interest. The finding that reading times for target sentences are facilitated by a preceding prime suggests that comprehenders are able to build representations of ungrammatical Subject islands and then

draw on those representations to facilitate later processing of that same structure. In other words, Experiment 2 suggests that priming is possible in Subject islands. Moreover, the pattern of differences between Lag1 and Lag5 suggests that the type of priming observed for Subject islands may be attributed to rapid decay of lingering structural activation. This is similar to what has been proposed to account for short-term priming in grammatical sentences.

Conversely, reading times between prime-target pairs in CNPC islands did not appear to differ in Lag1. Despite results from the stop-being-grammatical task (see *Differences between CNPC and Subject Islands: The Stop Being Grammatical Task*), we find no reading time slowdowns associated with either the word signaling the filled-gap (w5) or the point where the processor recognizes the illicit extraction (w7) when sentences were close together. Surprisingly, we did observe a localized one-word priming effect (w5) for CNPC islands when primes and targets were far apart (i.e., in Lag5).

The reading time pattern presented by CNPC islands is difficult to interpret because no prior work has predicted a structural priming effect that *only* surfaces at longer intervals (Lag5) between prime and target. Even implicit learning accounts



of priming, which predict a long-lasting effect, do not do so *in the absence* of short-term ones. Moreover, reading times for CNPC islands did not behave as one might have expected based on the stop-being-grammatical task. Results from the stop-being-grammatical task (see *Differences between CNPC and Subject Islands: The Stop Being Grammatical Task*) showed that comprehenders begin perceiving CNPC islands to be ungrammatical as early as the fifth word in the sentence (with most comprehenders concurring by the seventh word). Thus, comprehenders seem aware of the ungrammaticality of CNPC islands relatively early in the sentence. Yet, we do not detect processing difficulty (reading time slowdowns) at any point in CNPC sentences when prime and target are close together (Lag1).

It is worth noting that the reading time patterns we found for CNPC islands *do* resemble those reported for this same island type by Sag et al. (2007) and Hofmeister and Sag (2010). They investigated different issues, but used the same self-paced reading paradigm and found that reading times for CNPC islands did not differ from those in fully grammatical sentences. Crucially, their results showed that manipulating a single processing-related factor (bare *wh*-phrases vs. *which*-phrases, see *Syntactic Satiation*) was sufficient to effectively produce a reading-time ‘floor effect’ in CNPC islands. Though it may be possible that reading times for CNPC islands in Experiment 2 also exhibited a similar floor effect, this account provides little explanation for why reading times slowdowns were not detected for CNPC *primes*, which are not facilitated by prior exposure. At the moment, we leave the question of why CNPC islands did not show expected reading time slowdowns as a question for future work.

In sum, Experiment 2 leads us to conclude the following: First, reading time facilitation effects from primes to targets in Subject islands suggest that comprehenders are able to build a syntactic structure for this purportedly ungrammatical island-violation structure in real time, and that this structure can facilitate subsequent processing. Second, the results for CNPC islands suggest that structure-building for island sentences may be limited: If, following Ivanova et al. (2012a,b, 2017), we treat processing facilitation as a diagnostic for structure-building, our results indicate that comprehenders only build structures for some ungrammatical sentences. Thus, the different patterns of priming observed for Subject vs. CNPC islands reinforce the idea that the mechanisms involved in facilitating comprehension of ungrammatical sentences may not be a uniform, across-the-board phenomenon. Third, our results suggest that the proximity between prime and target sentences can affect online processing of Subject and CNPC islands, though the effect manifests itself differently for the two island types.<sup>16</sup>

## GENERAL DISCUSSION

The goal of this work was to investigate the extent to which syntactic satiation (exposure-induced rating improvements in ungrammatical sentences) could be linked to syntactic priming

(processing facilitation as a consequence of prior exposure). We focused on two types of island structures – Complex-NP Constraint (CNPC) and Subject islands. Our work departed from traditional approaches in satiation, where rating improvements are compared over the entire course of the study, and instead focuses on improvements between exposure-to-exposure pairs (i.e., primes vs. targets). This type of comparison allowed us to investigate whether factors known to affect online sentence processing, such as proximity of exposure and (less reliably) lexical repetition, could affect *judgments* of sentences similarly. If so, it may be possible to link priming and satiation to similar underlying mechanisms. Experiment 1 found that ratings for CNPC islands were improved by a preceding CNPC prime but only when primes and targets were separated by only one intervening sentence; when prime and target sentences were separated by five interveners, this effect was no longer detected. Subject islands, by contrast, saw no rating improvement either when prime-target pairs were close together, or when they were further apart. We further probed differences between CNPC and Subject island using the stop-being-grammatical task (see *Differences between CNPC and Subject Islands: The Stop Being Grammatical Task*). These results showed that differences between island types were not due to superficial differences in the position of the *wh*-gap (sentence-finally in CNPC vs. immediately after the head of the island phrase in Subject islands).

Given the results of Experiment 1, we then asked whether rating improvements simply reflected end-of-sentence, meta-linguistic judgment processes or whether they reflected *online* incremental comprehension processes for ungrammatical sentences. To do this, we used an online metric, reading time, to tap into structure-building and processing facilitation during the course of ungrammatical sentence comprehension. In Experiment 2, Subject islands showed reading times improvements over several words in our region of interest when primes and targets were close together (Lag1). However, when sentences were further apart (Lag5), these improvements persisted over only a single word in the region of interest. We also found that reading times for CNPC islands did not differ from each other in Lag1, suggesting that seeing one CNPC island did not facilitate CNPC processing when sentences were close together. But, when CNPC sentences were further apart, we did detect a (unexpected) single-word priming effect for CNPC islands such that target sentences were read slower than their prime counterparts.

Crucially, our results revealed a disjunction between Experiment 1 (acceptability ratings) and Experiment 2 (reading times) for both Subject and CNPC islands: Though we found no prime-to-target rating improvements for Subject islands in Experiment 1, we did find facilitated reading times in Experiment 2. This suggests that the processing of Subject islands can be facilitated (i.e., primed) by prior exposure during *online* comprehension, but that facilitation may not be sufficiently powerful to spill over to participants’ end-of-sentence offline acceptability ratings (see also Phillips, 2013 for a discussion of processing difficulty vs. well-formedness).

Meanwhile, CNPC islands did show prime-to-target rating improvements from a local exposure in Experiment 1, but

<sup>16</sup> Again, numerical differences between Lag1 and Lag5 were not compared directly, as discussed in footnote 7.

those improvements did not correspond to *online* reading time/processing improvements in Experiment 2. The lack of reading-time priming effects in CNPC islands may suggest that comprehenders do not construct a *syntactic* representation for CNPC islands in real time. Instead, we suggest that the acceptability rating improvements we observed with CNPC islands may be attributable not to *structural* priming, but to a different type of adaptation by the processor. For example, prior work on the processing of ungrammatical sentences has shown that there are many non-syntactic alternatives – based on frequency (e.g., Hare et al., 2003), discourse context (e.g., Spivey-Knowlton et al., 1993), plausibility (e.g., Ferreira, 2003), and simple word order heuristics (e.g., Ferreira, 2003) – through which comprehenders might choose to interpret an anomalous structure (see Pickering and van Gompel, 2006 for review). If alternative routes are more accessible than the syntactic structure-building route when comprehenders encounter a CNPC island, they will presumably opt for a non-syntactic approach. Thus, our failure to detect online facilitation effects with CNPC islands may be related to the viability of a non-structural processing route. Further research is needed to investigate this more directly. Under this view, the reading time slowdowns that we detected in the Lag5 group for CNPC islands hint that facilitation effects – even when not *structurally* driven – may be sensitive to the distance between exposures.

Taken together, our work points to some links between satiation (improvements in acceptability) and priming (facilitation in processing). First, we find that priming – and by extension, structure building – may be possible in Subject islands. And, while online processing effects were not reflected in end-of-sentence rating improvements, the presence of an online facilitation effect suggests that we cannot rule out the possibility of priming in ungrammatical sentences. Further, improvements observed for Subject and CNPC islands appear to be sensitive to the distance between prime and target sentences. Specifically, improvements – in terms of ratings (Experiment 1) or reading times (Experiment 2) – that emerged as a result of prior exposure were present when sentences are close together (Lag1), but absent when exposures are further apart (Lag5). One possibility, then, may be that both satiation and priming are linked to a short-term mechanism such as residual activation of structural representations that decay rapidly. Importantly, our results do not suggest that satiation should simply be equated with priming. While some of the results here may be compatible with ‘satiation as priming,’ it is premature at this stage to equate the two without further investigating factors such as the role of lexical repetition, (the absence of) long-term priming effects, etc.

## Implications for Theories of Island Constraints

Prior work has sought to directly address which factors might contribute to the different patterns of satiation across island types (cf. Hiramatsu, 2000; Kluender, 2004; Sag et al., 2007; Crawford, 2012; Chaves and Dery, 2014; *inter alia*). That issue is not the main focus of the experiments reported in this paper. However, both Experiments 1 and 2 suggest that Subject islands and CNPC

islands behave differently. Therefore, it may be reasonable to suggest that what has been grouped under the same ‘satiation’ umbrella may actually be two different underlying mechanisms, targeting different kinds of island violations, that happen to yield superficially similar consequences.

Prior work has attempted to classify island constraints under different syntactic (e.g., Ross, 1967; Huang, 1982; Chomsky, 1986; Rizzi, 1990) or semantic (e.g., Szabolcsi and Zwarts, 1993) mechanisms. To date, though, these typologies (e.g., “strong” vs. “weak” island effects) are neither very straight-forward nor fully agreed-upon (Szabolcsi and den Dikken, 2003; Szabolcsi, 2006; etc.). However, theories (Ross, 1967; Kluender, 1998, 2004; Hiramatsu, 2000; etc.) that suggest a typological distinction between CNPC and Subject islands may be able to capture the pattern of results presented here. For instance, some accounts consider CNPC islands to be “weak” and Subject islands to be “strong” by virtue of the severity of the violation (quantified in terms of subadjacency violations).<sup>17</sup> Though our work cannot speak to the validity of these classifications, it is worth noting that our results do provide evidence against grouping CNPC and Subject islands as a natural class. Clearly, further work is required to pinpoint what precisely defines the asymmetric satiation and priming effects that we observe.

The different pattern of behaviors for CNPC and Subject islands may also speak to an ongoing debate concerning the status of island violations in general. On one hand, with CNPC islands we (unexpectedly) found reading time differences between primes and targets when primes and targets were far apart but not when they were close together. This could be argued to lend support to accounts that primarily attribute island effects to processing effects (e.g., Kluender and Kutas, 1993; Kluender, 1998, 2004; Sag et al., 2007; Hofmeister and Sag, 2010; Pearl and Sprouse, 2012, 2015; but see Phillips, 2013). On the other hand, online facilitation effects for Subject islands were not strong enough to ‘spill over’ to acceptability improvements. This suggests that while the acceptability of island sentences may be affected by processing-related factors, attempts to locate island effects wholly outside the grammar are insufficient (Ross, 1967; Chomsky, 1986; Sprouse et al., 2012a,b; Phillips, 2013; Yoshida et al., 2014). As in the case of satiation, it may be that the role of processing-related factors may affect these two island types differently.

## Implications for Methodology

Traditional measures of satiation have relied on acceptability judgments, which is a consequence of how satiation as a phenomenon has been defined. However, our results show that there is a benefit to looking at satiation using multiple methods.

<sup>17</sup>In other accounts, both CNPC and Subject islands are considered “strong” islands; but, these accounts cannot explain the difference between island types observed here. We, therefore, use the terminology “weak” and “strong” here simply to follow the convention that was used by the relevant work (Kluender, 1998, 2004; Hiramatsu, 2000; etc.). Though, as noted above, the distinction between “weak” and “strong” islands is not straightforward and still an open question (Szabolcsi and den Dikken, 2003; Szabolcsi, 2006; etc.). What is critically relevant is that – regardless of terminology – prior work which has independently suggested a distinction between CNPC and Subject islands has the potential to account for differences observed here.

Ratings from the acceptability judgment task (Experiment 1) provide a ‘first look’ into the potential link between satiation and priming. Strikingly, once we adapted the task to an online measure (Experiment 2), it became apparent that acceptability ratings alone did not allow us to fully differentiate between the mechanisms targeting the two different sentence types. The emerging picture is admittedly complex, but adds new empirical evidence to a subfield of linguistics – satiation research – that has been characterized by a lack of consensus from the outset.

Finally, while prime-target proximity effects have been thoroughly investigated in the priming literature, our work is the first (to our knowledge) to take some initial steps toward investigating proximity in studies of acceptability ratings. Therefore, an independent contribution of our work is to highlight the need to control for distance between targets in acceptability judgment tasks.

## ETHICS STATEMENT

All studies reported in this paper were reviewed and approved by the University of Southern California University Park Institutional Review Board, which is fully accredited by the Association for the Accreditation of Human Research Protection

## REFERENCES

- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bernolet, S., Hartsuiker, R. J., and Pickering, M. J. (2013). From language-specific to shared syntactic representations: the influence of second language proficiency on syntactic sharing in bilinguals. *Cognition* 127, 287–306. doi: 10.1016/j.cognition.2013.02.005
- Bock, K. (1986). Syntactic persistence in language production. *Cogn. Psychol.* 18, 355–387. doi: 10.1016/0010-0285(86)90004-6
- Bock, K., Dell, G. S., Chang, F., and Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition* 104, 437–458. doi: 10.1016/j.cognition.2006.07.003
- Bock, K., and Griffin, Z. M. (2000). The persistence of structural priming: transient activation or implicit learning? *J. Exp. Psychol. Gen.* 129, 177–192. doi: 10.1037/0096-3445.129.2.177
- Boland, J. E., Tanenhaus, M. K., and Garnsey, S. M. (1990). Evidence for the immediate use of verb control information in sentence processing. *J. Mem. Lang.* 29, 413–432. doi: 10.1016/0749-596X(90)90064-7
- Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., and Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: evidence from *Wh*-questions. *J. Mem. Lang.* 34, 774–806. doi: 10.1006/jmla.1995.1034
- Branigan, H. P. (2007). Syntactic priming. *Lang. Linguist. Compass* 1, 1–17. doi: 10.1111/j.1749-818X.2006.00001.x
- Branigan, H. P., Pickering, M. J., and Cleland, A. A. (1999). Syntactic priming in written production: evidence for rapid decay. *Psychon. Bull. Rev.* 6, 535–540. doi: 10.3758/BF03212972
- Branigan, H. P., Pickering, M. J., Liversedge, S. P., Stewart, A. J., and Urbach, T. P. (1995). Syntactic priming: investigating the mental representation of language. *J. Psycholinguist. Res.* 24, 489–506. doi: 10.1007/BF02143163
- Chang, F., Dell, G. S., and Bock, K. (2006). Becoming syntactic. *Psychol. Rev.* 113, 234–272. doi: 10.1037/0033-295X.113.2.234

Programs (AAHRPP). Due to the nature of the experiments, the Institutional Review Board determined that written consent was not needed.

## AUTHOR CONTRIBUTIONS

MD and EK conceptualized and designed the experiments. MD acquired the data and conducted the statistical analyses. Both MD and EK interpreted the data and wrote the manuscript.

## ACKNOWLEDGMENTS

We thank Stefan Keine, Jon Sprouse, Andrew Simpson, Rand Wilcox, and the USC Sentence Processing Lab for their insight and feedback on this work. We also thank our reviewers for the extremely thoughtful and encouraging feedback that contributed to the improvement of this work. Finally, we thank the audiences at the 34th West Coast Conference on Formal Linguistics and the 22nd Architectures and Mechanisms for Language Processing Conference, where some of this work was presented. A portion of the analyses presented here also appeared in the Proceedings of the 34th West Coast Conference on Formal Linguistics.

- Chang, F., Dell, G. S., Bock, K., and Griffin, Z. M. (2000). Structural priming as implicit learning: a comparison of models of sentence production. *J. Psycholinguist. Res.* 29, 217–229. doi: 10.1023/A:1005101313330
- Chaves, R. P., and Dery, J. E. (2014). “Which subject islands will the acceptability of improve with repeated exposure?,” in *Proceedings of the 31st West Coast Conference on Formal Linguistics*, ed. R. E. Santana-LaBarge (Somerville, MA: Cascadilla Proceedings Project), 96–106.
- Chomsky, N. (1986). *Barriers*. Cambridge, MA: MIT Press.
- Cinque, G. (1990). *Types of A'-dependencies*. Cambridge, MA: MIT Press.
- Cleland, A. A., and Pickering, M. J. (2003). The use of lexical and syntactic information in language production: evidence from the priming of noun phrase structure. *J. Mem. Lang.* 49, 214–230. doi: 10.1016/S0749-596X(03)00060-3
- Crain, S., and Fodor, J. D. (1985). “How can grammars help parsers?,” in *Natural Language Parsing: Psycholinguistic, Computational, and Theoretical Perspectives*, eds D. R. Dowty, L. Karttunen, and A. M. Zwicky (Cambridge: Cambridge University Press), 94–128. doi: 10.1017/CBO9780511597855.004
- Crawford, J. (2012). “Using syntactic satiation to investigate subject islands,” in *Proceedings of the 29th West Coast Conference on Formal Linguistics*, eds J. Choi, E. A. Hogue, J. Punske, D. Tat, J. Schertz, and A. Trueman (Somerville, MA: Cascadilla Proceedings Project), 38–45.
- Davies, W. D., and Dubinsky, S. W. (2003). On extraction from NPs. *Nat. Lang. Linguist. Theory* 21, 1–37. doi: 10.1023/A:1021891610437
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cogn. Psychol.* 47, 164–203. doi: 10.1016/S0010-0285(03)00005-7
- Francom, J. C. (2009). *Experimental Syntax: Exploring the Effect of Repeated Exposure to Anomalous Syntactic Structure – Evidence from Rating and Reading Tasks*. Ph.D. dissertation, University of Arizona, Tucson, AZ.
- Frazier, L., and Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Lang. Cogn. Process.* 4, 93–126. doi: 10.1080/01690968908406359
- Gibson, E., and Hickok, G. (1993). Sentence processing with empty categories. *Lang. Cogn. Process.* 8, 147–161. doi: 10.1080/01690969308406952

- Goodall, G. (2011). Syntactic satiation and the inversion effect in English and Spanish *Wh*-questions. *Syntax* 14, 29–47. doi: 10.1111/j.1467-9612.2010.00148.x
- Goodall, G. (2015). The D-linking effect on extraction from islands and non-islands. *Front. Psychol.* 5:1493. doi: 10.3389/fpsyg.2014.01493
- Hare, M., McRae, K., and Elman, J. L. (2003). Sense and structure: meaning as a determinant of verb subcategorization preferences. *J. Mem. Lang.* 48, 281–303. doi: 10.1016/S0749-596X(02)00516-8
- Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., and Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays. *J. Mem. Lang.* 58, 214–238. doi: 10.1016/j.jml.2007.07.003
- Hartsuiker, R. J., and Kolk, H. H. J. (1998). Syntactic persistence in Dutch. *Lang. Speech* 41, 143–184. doi: 10.1177/002383099804100202
- Hartsuiker, R. J., Pickering, M. J., and Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychol. Sci.* 15, 409–414. doi: 10.1111/j.0956-7976.2004.00693.x
- Hiramatsu, K. (2000). *Assessing Linguistic Competence: Evidence from Children's and Adults' Acceptability Judgments*. Ph.D. dissertation, University of Connecticut, Storrs, CT.
- Hofmeister, P., and Sag, I. A. (2010). Cognitive constraints and island effects. *Language* 86, 366–415. doi: 10.1353/lan.0.0223
- Hofmeister, P., Staua Casasanto, L., and Sag, I. A. (2012a). How do individual cognitive differences relate to acceptability judgments?: a reply to Sprouse, Wagers, and Phillips. *Language* 88, 390–400. doi: 10.1353/lan.2012.0025
- Hofmeister, P., Staua Casasanto, L., and Sag, I. A. (2012b). Misapplying working-memory tests: a reductio ad absurdum. *Language* 88, 408–409. doi: 10.1353/lan.2012.0033
- Huang, C.-T. J. (1982). *Logical Relations in Chinese and the Theory of Grammar*. Ph.D. thesis Dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- Ivanova, I., Branigan, H. P., McLean, J. F., Costa, A., and Pickering, M. J. (2017). Do you what I say? People reconstruct the syntax of anomalous utterances. *Lang. Cogn. Neurosci.* 32, 175–189. doi: 10.1080/23273798.2016.1236976
- Ivanova, I., Pickering, M. J., Branigan, H. P., McLean, J. F., and Costa, A. (2012a). The comprehension of anomalous sentences: evidence from structural priming. *Cognition* 122, 193–209. doi: 10.1016/j.cognition.2011.10.013
- Ivanova, I., Pickering, M. J., McLean, J. F., Costa, A., and Branigan, H. P. (2012b). How do people produce ungrammatical utterances? *J. Mem. Lang.* 67, 355–370. doi: 10.1080/02643294.2012.712957
- Karttunen, L. (1977). Syntax and semantics of questions. *Linguist. Philos.* 1, 3–44. doi: 10.1007/BF00351935
- Kaschak, M. P., and Glenberg, A. M. (2004). This construction needs learned. *J. Exp. Psychol. Gen.* 133, 450–467. doi: 10.1037/0096-3445.133.3.450
- Kluender, R. (1998). “On the distinction between strong and weak islands: a processing perspective,” in *Syntax and Semantics: The Limits of Syntax*, Vol. 29, eds P. Culicover and L. McNally (San Diego, CA: Academic Press), 241–279.
- Kluender, R. (2004). “Are subject islands subject to a processing account?” in *Proceedings of the West Coast Conference on Formal Linguistics* 23, eds V. Chand, A. Kelleher, A. Rodriguez, and B. Schmeier (Somerville, MA: Cascadilla Press), 101–125.
- Kluender, R., and Kutas, M. (1993). Subjacency as a processing phenomenon. *Lang. Cogn. Process.* 8, 573–633. doi: 10.1080/01690969308407588
- Luka, B., and Barsalou, L. (2005). Structural facilitation: mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *J. Mem. Lang.* 52, 436–459. doi: 10.1016/j.jml.2005.01.013
- Maling, J., and Zaenen, A. (1982). “A phrase structure account of Scandinavian extraction phenomena,” in *The Nature of Syntactic Representation*, eds P. Jacobson and G. Pullum (Dordrecht: Riedel), 229–282.
- Pearl, L., and Sprouse, J. (2012). “Computational models of acquisition for islands,” in *Experimental Syntax and Islands Effects*, eds J. Sprouse and N. Hornstein (Cambridge: Cambridge University Press), 109–131.
- Pearl, L., and Sprouse, J. (2015). Computational modeling for language acquisition: a tutorial with syntactic islands. *J. Speech Lang. Hear. Res.* 58, 740–753. doi: 10.1044/2015\_JSLHR-L-14-0362
- Pesetsky, D. (1987). “*WH*-in-situ: movement and unselective binding,” in *The Representation of (In)definiteness*, eds E. Reuland and A. ter Muelen (Cambridge, MA: MIT Press), 98–129.
- Pesetsky, D. (2000). *Phrasal Movement and its Kin*. Cambridge, MA: MIT Press.
- Phillips, C. (2006). The real-time status of island constraints. *Language* 82, 795–823. doi: 10.1353/lan.2006.0217
- Phillips, C. (2013). “On the nature of island constraints. I: language processing and reductionist accounts,” in *Experimental Syntax and Island Effects*, eds J. Sprouse and N. Hornstein (Cambridge: Cambridge University Press), 64–108.
- Pickering, M. J., and Branigan, H. P. (1998). The representation of verbs: evidence from syntactic priming in language production. *J. Mem. Lang.* 39, 633–651. doi: 10.1006/jmla.1998.2592
- Pickering, M. J., Branigan, H. P., Cleland, A. A., and Stewart, A. J. (2000). Activation of syntactic information during language production. *J. Psycholinguist. Res.* 29, 205–216. doi: 10.1023/A:1005149129259
- Pickering, M. J., and van Gompel, R. P. G. (2006). “Syntactic Parsing,” in *Handbook of Psycholinguistics*, 2nd Edn. eds M. Traxler and M. A. Gernsbacher (Amsterdam: Academic Press), 455–488.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rizzi, L. (1990). *Relativized Minimality*. Cambridge, MA: MIT Press.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition* 42, 107–142. doi: 10.1016/0010-0277(92)90041-F
- Rohde, D. (2010). *Linger*. Available at: <http://tedlab.mit.edu/~dr/Linger/>
- Ross, J. R. (1967). *Constraints on Variables in Syntax*. Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- Sag, I. A., Hofmeister, P., and Snider, N. (2007). “Processing complexity in subjacency violations: the complex noun phrase constraint,” in *Proceedings of the 43rd Annual Meeting of the Chicago Linguistic Society*, eds M. Elliott, J. Kirby, O. Sawada, E. Staraki, and S. Yoon (Chicago, IL: CLS), 215–229.
- Scheepers, C. (2003). Syntactic priming of relative clause attachments: persistence of structural configuration in sentence production. *Cognition* 89, 179–205. doi: 10.1016/S0010-0277(03)00119-7
- Snyder, W. (2000). An experimental investigation of syntactic satiation effects. *Linguist. Inq.* 31, 575–582. doi: 10.1162/002438900554479
- Snyder, W. (2017). *On the Nature of Syntactic Satiation*. Storrs, CT: University of Connecticut.
- Spivey-Knowlton, M. J., Trueswell, J. C., and Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: discourse and semantic influences in parsing reduced relative clauses. *Can. J. Exp. Psychol.* 37, 276–309. doi: 10.1037/h0078826
- Sprouse, J. (2007). Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1, 123–134.
- Sprouse, J. (2009). Revisiting Satiation: evidence for an equalization response strategy. *Linguist. Inq.* 40, 329–341. doi: 10.1162/ling.2009.40.2.329
- Sprouse, J., Wagers, M., and Phillips, C. (2012a). A test of the relation between working-memory capacity and syntactic island effects. *Language* 88, 82–123. doi: 10.1353/lan.2012.0004
- Sprouse, J., Wagers, M., and Phillips, C. (2012b). Working-memory capacity and island effects: a reminder of the issues and the facts. *Language* 88, 401–407. doi: 10.1353/lan.2012.0029
- Szabolcsi, A. (2006). “Strong vs. weak islands,” in *The Blackwell Companion to Syntax*, Vol. 4, eds M. Everaert and H. van Riemsdijk (Malden, MA: Blackwell), 479–531. doi: 10.1002/9780470996591.ch64
- Szabolcsi, A., and den Dikken, M. (2003). “Islands,” in *The Second Glot International State-of-the-Article Book*, eds L. Cheng and R. Sybesma (Berlin: Mouton de Gruyter), 213–240.
- Szabolcsi, A., and Zwarts, F. (1993). Weak islands and an algebraic semantics for scope taking. *Nat. Lang. Semant.* 1, 235–284. doi: 10.1007/BF00263545
- Thothathiri, M., and Snedeker, J. (2008a). Give and take: syntactic priming during spoken language comprehension. *Cognition* 108, 51–68. doi: 10.1016/j.cognition.2007.12.012
- Thothathiri, M., and Snedeker, J. (2008b). Syntactic priming during language comprehension in three and four year old children. *J. Mem. Lang.* 58, 188–213. doi: 10.1016/j.jml.2007.06.012
- Tooley, K. M., and Traxler, M. J. (2010). Syntactic priming effects in comprehension: a critical review. *Lang. Linguist. Compass* 4, 925–937. doi: 10.1111/j.1749-818X.2010.00249.x
- Traxler, M. J. (2008). Lexically independent syntactic priming of adjunct relations in on-line sentence comprehension. *Psychon. Bull. Rev.* 15, 145–155. doi: 10.3758/PBR.15.1.149

Yoshida, M., Kazanina, N., Pablos, L., and Sturt, P. (2014). On the origin of islands. *Lang. Cogn. Neurosci.* 29, 761–770. doi: 10.1080/01690965.2013.788196

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2017 Do and Kaiser. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# On the Diversity of Linguistic Data and the Integration of the Language Sciences

Roberta D'Alessandro<sup>1\*</sup> and Marc van Oostendorp<sup>2,3</sup>

<sup>1</sup> Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, Netherlands, <sup>2</sup> Meertens Instituut (KNAW), Amsterdam, Netherlands, <sup>3</sup> Centre for Language Studies, Radboud University Nijmegen, Nijmegen, Netherlands

An integrated science of language is usually advocated as a step forward for linguistic research. In this paper, we maintain that integration of this sort is premature, and cannot take place before we identify a common object of study. We advocate instead a science of language that is inherently multi-faceted, and takes into account the different viewpoints as well as the different definitions of the object of study. We also advocate the use of different data sources, which, if non-contradictory, can provide more solid evidence for linguistic analysis. Last, we argue that generative grammar is an important tile in the puzzle.

## OPEN ACCESS

### Edited by:

Ángel J. Gallego,  
Universitat Autònoma de Barcelona,  
Spain

### Reviewed by:

Luis Lopez,  
University of Illinois at Chicago,  
United States  
Javi Fernández Sánchez,  
University of Gdansk, Poland

### \*Correspondence:

Roberta D'Alessandro  
r.dalessandro@uu.nl

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 23 August 2017

**Accepted:** 01 November 2017

**Published:** 23 November 2017

### Citation:

D'Alessandro R and  
van Oostendorp M (2017) On  
the Diversity of Linguistic Data  
and the Integration of the Language  
Sciences. *Front. Psychol.* 8:2002.  
doi: 10.3389/fpsyg.2017.02002

**Keywords:** generative grammar, minimalism, biolinguistics, construction grammar, functionalism, cognition, linguistic data

## INTRODUCTION

In a recent article, Christiansen and Chater (2017) (henceforth CC) argue in favor of an 'integrated science of language.' Just as "integration and interaction between levels of analysis and diverse data is ubiquitous [in] the physical and biological sciences," progress in linguistics can only be guaranteed by taking into account a wide variety of data from a range of different sources.

We suspect there are not many linguists who would disagree with the observation that attempts to integrate knowledge and to facilitate interaction between students of language working at different 'levels of analysis' would probably be beneficial to the field. Clearly, the number and variety of empirical sources that have become available in recent decades for anyone interested in the topic of human language has broadened considerably, and continues to do so: from ultrasound measurements to automatic exploration of large amounts of words used on social media, and from fieldwork notes on Amazonian languages that are already extinct to neurolinguistics data on people learning artificial languages while in an MRI machine – all of these can potentially shed light on the question what human language is and how it works. It is regrettable indeed that the boundaries between the people studying all these different types of data are seldom crossed.

CC, however, see one major obstacle in this integration: 'Chomskyan' linguistics. They state: "Many of the phenomena that have become the focus of syntactic theory are so abstract that they are often difficult to connect even with specific linguistic phenomena, let alone with experiments on how people process language or observations of how children learn their native tongue." For this reason, they propose replacing generative grammar with construction grammars (for which they cite Goldberg, 2006; strangely, they do not cite *any* reference for generative grammar), because their "quasi-regular nature [. . .] allows them to capture both the rule-like patterns as well as the myriad of exceptions that often are excluded by fiat from the old view built on abstract rules."

They do not give precise details about how construction grammar makes better predictions than generative grammar.

The structure of CC's argument is very similar to that put forward by Levinson and Evans (2010) (henceforth LE), although CC do not mention that earlier paper. LE state that "[generativists] draw on a very small subset of the data – especially, intuitions about complex clauses. Meanwhile, the available data types (corpora, typological databases, multimedia records), and the range of data over the languages of the world, has vastly increased in recent years, as has the scientific treatment of grammatical intuitions" and they contrast this with "the vastly increased quantity, quality and types of data now available to the descriptive and comparative linguist." Like CC, LE seem to argue for an integrated science of language, in which everybody is welcome to contribute, except for the Chomskians.

We believe that CC and LE misrepresent the range of methodologies that are used by scholars sympathetic to the generative paradigm, in which many kinds of data have also been studied recently, and sometimes with considerable success. We agree with them that the question of how the body of ideas that constitutes generative grammar should relate to the wealth of data that is available to us is important, as is whether there is any place for generative inquiry/biolinguistics (Jenkins, 2000) in an integrated science of language. We want to discuss both of these questions in this short contribution.

## THE ONTOLOGY OF LANGUAGE IN GENERATIVE GRAMMAR

Anybody who seriously aims to undertake an integrated study of language should first note that there is very little agreement about the ontology of the object of study among linguists. One clear opposition is that which could be referred to as Chomsky vs. Saussure. In the first line of thought, language is seen as a cognitive object, something which resides in the mind of an individual speaker (Chomsky, 1957, 1965 *ff.*), and communities present chaotic mixtures of these idiolects. The other line is the Saussurean view (also foundational to, e.g., Labovian linguistics) in which language resides in a community, and the language production of individual speakers is an imperfect reflection of those speakers. Both of these positions seem coherent in their own right, and work from both schools can be combined, although they obviously conflict in their ultimate vision of what language is. There are also other visions available, such as the Platonic view (Postal, 2009) which sees language as "a purely abstract object, on a par with those of mathematics."

It is important to point out that such approaches are not easily reconciled, as they seem incommensurable in the well-known sense of Kuhn (1962): they are different in scope. This does not mean that data or even insights cannot be transferred from one to the other; witness successful work that has been done over the years that shows otherwise (see for instance Kroch's, 1994; Cornips and Corrigan's, 2005; and Adger's, 2016 work on "socio-syntax," to use Adger's term). Such interactions are, however, more complicated than different 'levels of analysis' (say, the subatomic level to the atomic level) in physics; the linguistic

disciplines are simply not easily integrated in any reasonable sense of that word.

It is not clear where CC and LE stand in this debate about the ontology of language. On the one hand, there is a certain sympathy in both papers for so-called cognitive grammar (of which construction grammar is usually seen as a variant, i.e., *Cognitive Construction Grammar*, inspired by Goldberg, 1995 *ff.*), although both papers occasionally refer to 'culture' and 'communication' as sources of explanation, leaving open the question of how these different modalities relate to each other (whether they are to be seen as 'different levels of analysis'). At first sight, the first victim of a revolutionary 'integration' along the lines of LE and CC seems to be the Saussurian/Labovian view of language rather than the Chomskian view. In any case, there seems to be no attempt to reconcile these different views with one another, or with the Platonic view (but see Watumull, 2013 on the potential compatibility of Platonism and biolinguistics).

CC make use of a very salient metaphor: language is like a crossword, where figuring out one clue will help figure out the next clue. They describe the way that language acquisition takes place in a crossword-like fashion. Children are sensitive to "multiple sources of probabilistic information available in the linguistic input: from the sound of words to their co-occurrence patterns to information from semantic and pragmatic contexts." According to CC, there is no need to postulate an innate set of pre-existing categories, for instance: children can infer categories from statistical analyses of distribution. The construction grammar approach accounts very well, CC maintain, for the diversity of the world's languages.

The first observation that comes to mind is that this view of generative grammar is inaccurate: many generative approaches do not postulate pre-existing categories (see the work of Wiltschko or Biberauer on emergentist features). Then, it seems to us that construction grammar lacks predictive power: much like the old transformational grammar rules, in construction grammar everything goes, as long as there is evidence for it. No restriction is imposed on structures because of the system itself. We know that this is not accurate. Although many of the macro-parametric approaches have proved unsuccessful, some generalizations on co-occurring structural properties across languages cannot be easily denied.

Keeping the empirical coverage aside for the moment, we submit that, using CC's metaphor, integration is impossible, because the clues are not for the same crossword. It is possible that convincing theories will be developed in which a link can be found between the psychological and the sociological, and between each of these and the abstract, in which case we could hope to build a truly integrative framework for the language sciences. None of this means that one particular view (of those mentioned) on this issue on this is inherently superior. As Chomsky (2001:34) phrased it:

Internalist biolinguistic inquiry [Chomsky's term for what we call Chomskian linguistics here] does not, of course, question the legitimacy of other approaches to language, any more than internalist inquiry into bee communication invalidates the study of how the relevant

internal organization of bees enters into their social structure. The investigations do not conflict; they are mutually supportive. In the case of humans, though not other organisms, the issues are subject to controversy, often impassioned, and needless.

It should be added that Chomsky's practice or that of his followers may not always have conformed to this dictum, and have sometimes suggested that the *only* way of doing linguistics is by doing generative grammar, or that 'language' is a synonym for 'the innate capacity to acquire language.'

We propose, then, that rather than attempting a premature integration of different branches of linguistics, we should maximally profit from the *mosaic* nature of the field: the many different viewpoints that are taken on subject matters that have many things in common. Integration, as proposed in CC and LE, would lead to severe impoverishment of those points of view, forcing all linguistics to work in one frame (construction grammar) that was never designed to answer all questions and that has not had the time to be sufficiently tested. To borrow another set of terms from Kuhn, it is as if CC and LE want to move immediately from a period of (perceived) crisis to normal science, without wanting to go through the stage of paradigm shift. We think linguistics is not yet ready to be a coherent normal science, and it would be detrimental to pretend that it is: one can obviously always carry out numerous 'empirical studies,' but without a solid base it is impossible to achieve the kind of cumulative effect that is so typical of 'real science.'

Generative grammar, or more precisely a form of biolinguistics, based on a view in which language is primarily an internal tool for thought or expression of thought, cannot be excluded from such a multifaceted way of studying language. One can argue, if one sees reasons to do so, that current work on this matter is not satisfactory or is even wrong, but one cannot *a priori* deny that there are reasons to engage in such an enterprise.

A mosaic view on linguistics, we find, is a better metaphor than a crossword: we have tiles of different shapes, different colors, and differing importance. Inserting one tile in the mosaic will only give us a clue about what comes next, what is adjacent. Only the combination of all tiles allows us to see the full picture. If some tiles are missing, we will be able to figure them out. But, importantly, tiles do not resemble crossword clues, as they are not uniform in nature. Insights from different disciplines can all contribute tiles. The combination of all these tiles, including those regarding structural dependencies coming from generative grammar, will give us a picture of language.

## THE DATA FOR GENERATIVE GRAMMAR

This, then, seems to us the most reasonable position for generative grammar among the language sciences: as an approach to understanding what is specific about human language (in particular syntax) and to specifying what computational capacity the human mind needs to be able to acquire and use syntax. In no way should this prevent generative grammarians from collaborating with scholars working on other aspects, sometimes even within a completely different paradigm. We have already

mentioned above work on the crossroads with sociolinguistics above, but we should also consider work such as that by Andrea Moro on neurolinguistics, by George Walkden and David Lightfoot on diachronic linguistics, and by William Snyder, Maria Teresa Guasti, and Jason Rothman on psycholinguistics and acquisition.

It follows from this list that CC and LE's view of the range of types of data on which generative work is based is too pessimistic. There is also no reason why it could not widen more. For instance, the fact that intuitions often lack a quantitative component does not make them inherently less valuable, as Labov (1987), one of the fathers of quantitative linguistics, reminds us:

But the qualitative is not easily displaced. Many forms of linguistic behavior are categorically invariant. Furthermore, the number, variety and complexity of linguistic relations are very great, and it is not likely that a large proportion can be investigated by quantitative means. At present, we do not know the correct balance between the two modes of analysis.

On the contrary, any kind of scientific enterprise can only benefit from including as much empirical evidence as possible. As the eventual goal of generative grammar is to discover properties of the human mind, there is no such thing as direct evidence for this; there is no golden path. Intuitions have the advantage of being cheap and easy to acquire, but since they have their own inherent problems (they are not always as clear as we would want them to be; there can easily be interference with external norms on language, etc.), it seems that extending the empirical basis can only be a good thing.

For this we could follow, for instance, the taxonomy offered in van Oostendorp (2013), which was made for phonology, but can be easily extended to syntax: this taxonomy recognizes four types of evidence: traditional evidence (such as judgments, or the Wug tests); experimental evidence (such as that acquired in psycholinguistic or neurolinguistics laboratories); evidence from large databases and corpora (whether found in historical archives or tagged collections of modern text); and formal evidence (the results of computer modeling, analysis of formal elegance, etc.). All of these general types of data can be helpful beyond what we can establish from judgments alone. For instance, artificial language learning experiments (Moro, 2016) have shown that 'crazy patterns,' predicted not to exist by current theories, involve a different part of the brain than 'realistic patterns.' Automatic searching of large corpora can lead us to find patterns that an analyst would never have thought of independently. Computer modeling helps to make theories maximally explicit and thereby exposes hidden flaws.

None of these data can give us direct access to what we are really interested in – an object of considerable abstractness. We can therefore only aim to find convergent evidence from many different sides. The work on these types of data can of course take place in cooperation with researchers with a slightly different focus, which can in fact improve the way we approach the object of study. It does not necessarily mean that one has to share the same view on what should be studied.



Finally, it should also be kept in mind that even people who consider themselves practitioners of Chomskyan generative syntax do not necessarily have the same interests. We feel that there is a rather wide consensus that there are at least two types: those working in some version of what used to be called Government and Binding Theory (Chomsky, 1981), taking an interest mostly in trying to explain patterns in individual language varieties; and those subscribing whole-heartedly to the Minimalist Program (Chomsky, 1995). The former will typically be closer to types of data such as those just listed, whereas for the latter, the analyses formed by G&B count as data of some kind. This is the kind of work that presumably led CC to their complaint that the analyses are “so abstract that they are often difficult to connect even with specific linguistic phenomena.” We hope to have shown by now that this vision is too narrow, as it presupposes that there is some non-theoretical way of deciding what “specific linguistic phenomena” are. However, all ‘linguistic phenomena’ are theory-laden and dependent on one’s ontology of language. Suggesting otherwise, and operating on the assumption that we have some pre-theoretical conception of the subject matter is, in our view, not going to lead linguistics very far.

## CONCLUSION

As sympathetic as it may sound at first sight, calls for ‘integration’ of the language sciences, such as those by CC and LE, do

## REFERENCES

- Adger, D. (2016). “Language variability in syntactic theory,” in *Rethinking Parameters*, ed. A. Mendikoetxea (Oxford: Oxford University Press). doi: 10.1093/acprof:oso/9780190461737.003.0002
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Hawthorne, NY: Walter De Gruyter.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge: MIT Press.
- Chomsky, N. (2001). “Derivation by phase,” in *Ken Hale: A Life in Language*, ed. M. Kenstowicz (Cambridge: MIT Press), 1–52.
- Christiansen, M. H., and Chater, N. (2017). Towards an integrated science of language. *Nat. Hum. Behav.* 1:0163. doi: 10.1038/s41562-017-0163
- Cornips, L., and Corrigan, K. (2005). “Towards an integrated approach to syntactic variation. A retrospective and prospective synopsis,” in *Syntax and Variation: Reconciling the Biological and the Social*, eds L. Cornips and K. Corrigan (Amsterdam: Benjamins).
- Goldberg, A. (1995). *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago, IL: University of Chicago Press.
- Goldberg, A. (2006). *Constructions at Work: The Nature of Generalization in Language*. New York, NY: Oxford University Press.
- Jenkins, L. (2000). *Biolinguistics: Exploring the Biology of Language*. Cambridge: Cambridge University Press.
- Kroch, A. (1994). “Morphosyntactic variation,” in *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory*, ed. K. Beals (Chicago, IL: CLS).

not take into account the fact that there is no consensus on what linguistics is about, or what the explananda are – and therefore what the data to be taken into account are. Rather than calling for an integration of this type, which in our view can only lead to multiple small case studies, and experiments without sufficient loopback to a strong theory, we think it is better to opt for a model of the language sciences as a mosaic of different views and methodologies, hoping that in this way – and by cooperating across the disciplines rather than dismissing some of them out of hand – we can achieve a better understanding of the multifaceted phenomenon that is human language.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon2020 research and innovation program (grant agreement No. 681959\_Microcontact).

- Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.
- Labov, W. (1987). *Some Observations on the Foundation of Linguistics*. Philadelphia, PA: University of Pennsylvania.
- Levinson, S. C., and Evans, N. (2010). Time for a sea-change in linguistics: response to comments on ‘The Myth of Language Universals’. *Lingua* 120, 2733–2758. doi: 10.1016/j.lingua.2010.08.001
- Moro, A. (2016). *Impossible Languages*. Cambridge: MIT Press.
- Postal, P. M. (2009). The incoherence of Chomsky’s ‘biolinguistic’ ontology. *Biolinguistics* 3, 104–123.
- van Oostendorp, M. (2013). A consumer’s guide to phonological evidence. *Nordlyd* 40, 1. doi: 10.7557/12.2477
- Watumull, J. (2013). Biolinguistics and Platonism: Contradictory or Consilient? *Biolinguistics* 7, 301–315.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a past collaboration with the author RD.

Copyright © 2017 D'Alessandro and van Oostendorp. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Sentence Repetition as a Tool for Screening Morphosyntactic Abilities of Bilectal Children with SLI

Elena Theodorou<sup>1\*</sup>, Maria Kambanaros<sup>1</sup> and Kleanthes K. Grohmann<sup>2,3</sup>

<sup>1</sup> Department of Rehabilitation Sciences, Cyprus University of Technology, Limassol, Cyprus, <sup>2</sup> Department of English Studies, University of Cyprus, Nicosia, Cyprus, <sup>3</sup> Cyprus Acquisition Team, Nicosia, Cyprus

## OPEN ACCESS

### Edited by:

Ángel J. Gallego,  
Universitat Autònoma de Barcelona,  
Spain

### Reviewed by:

Randi Martin,  
Rice University, United States  
F. Sayako Earle,  
University of Delaware, United States

### \*Correspondence:

Elena Theodorou  
eleni.theodorou@cut.ac.cy

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 27 April 2017

**Accepted:** 17 November 2017

**Published:** 06 December 2017

### Citation:

Theodorou E, Kambanaros M and  
Grohmann KK (2017) Sentence  
Repetition as a Tool for Screening  
Morphosyntactic Abilities of Bilectal  
Children with SLI.  
*Front. Psychol.* 8:2104.  
doi: 10.3389/fpsyg.2017.02104

The clinical significance of sentence repetition tasks (SRTs) for assessing children's language ability is well-recognized. SRT has been identified as a good clinical marker for children with (specific) language impairment as it shows high diagnostic accuracy levels. Furthermore, qualitative analysis of repetition samples can provide information to be used for intervention protocols. Despite the fact that SRT is a familiar task in assessment batteries across several languages, it has not yet been measured and validated in bilectal settings, such as Cypriot Greek, where the need for an accurate screening tool is urgent. The aims of the current study are three-fold. First, the performance of a group of (Cypriot) Greek-speaking children identified with SLI is evaluated using a SRT that elicits complex morphosyntactic structures. Second, the accuracy level of the SRT for the identification of SLI is explored. Third, a broad error analysis is carried out to examine and compare the morphosyntactic abilities of the participating children. A total of 38 children aged 5–9 years participated in this study: a clinical group of children with SLI ( $n = 16$ ) and a chronological age-matched control group ( $n = 22$ ). The ability of the children to repeat complex morphosyntactic structures was assessed using a SRT consisting of 24 sentences. The results showed that the SRT yielded significant differences in terms of poorer performance of children with SLI compared to typically developing peers. The diagnostic accuracy of the task was validated, since regression analysis showed that the task is sensitive and specific enough to identify children with SLI. Finally, qualitative differences between children with SLI and those with TLD regarding morphosyntactic abilities were detected. This study showed that a SRT that elicits morphosyntactically complex structures could be a potential clinical indicator for SLI in Cypriot Greek. The task has the potential to be used as a referral criterion in order to identify children whose language needs to be evaluated further. Implications for speech–language therapists and policy-makers are discussed.

**Keywords:** screening, clinical marker, referral criterion, bilectalism, Cypriot Greek

## INTRODUCTION

Identifying and diagnosing children with specific language impairment (SLI) is characterized internationally by both clinicians and researchers as an exceptional challenge. The principal goal of the present study is to determine whether a sentence repetition task (SRT), which includes different morphosyntactic structures, can serve as an accurate screening task, and as such as a

referral criterion, for the early identification of SLI in Cypriot Greek-speaking children. In the long term, this will ensure access to early and comprehensive assessment for individuals with SLI and their families. The study also aims to examine whether sentence repetition can yield differences between groups of language-impaired vs. non-impaired participants in terms of morphosyntactic errors.

Whilst language acquisition is one of the most robust, yet largely intrinsically driven, processes of early childhood (e.g., Lenneberg, 1967; Chomsky, 1986), not all children acquire language fully or even effortlessly. The term SLI is applied to children that exhibit a significant deficit in language ability and yet, display normal hearing, have non-verbal intelligence in the broad range of normal with no obvious signs of neurological damage or social-emotional deprivation (Leonard, 1998; Bishop, 2014). We acknowledge that there is no consensus regarding the criteria for classification and the related terminology (Bishop et al., 2016), but an in-depth discussion on this matter is beyond the scope of this paper; we will subsequently employ the term SLI, noting that the “S” part may be debatable. The description of deviant or inferior language ability in SLI is usually based on (i) characteristics of children’s spontaneous speech output and (ii) children’s performance on linguistic tasks tapping into different language components (such as morphology, phonology, syntax, semantics, and pragmatics as well as the lexicon). There is now increasing evidence to suggest that children with SLI can present with different patterns of impairment based on which modules of the language system are impaired or spared, hence the absence of homogeneity in the disorder (e.g., Leonard, 1998; van der Lely, 2003; Friedmann and Novogrodsky, 2008).

Sentence repetition (also referred to as “sentence recall and sentence imitation”) taps into an individual’s ability to repeat the exact wording of what was just heard. In the more recent past, research interest has turned to the diagnostic accuracy of the task. Studies have revealed that sentence repetition is a good psycholinguistic indicator of SLI in that consistently high diagnostic accuracy levels have been shown. For English, the observed positive correlation between sentence repetition with a number of language tests that are used widely, such as the Preschool Language Scale-3 (Boucher and Lewis, 1997), the Receptive and Expressive One Word Picture Vocabulary Test (Brownwell, 2000), and the Sentence Recall Subtest of the CELF (Wiig et al., 1992), has led to the assumption that the task can be a clinical marker for language impairment (Chiat and Roy, 2008). The term “clinical marker” refers to a particular structure that denotes SLI and for the purposes of this study it will be used for a task that includes different structures in accordance with similar research in the field (e.g., Conti-Ramsden et al., 2001; Stokes et al., 2006; Riches et al., 2010; Leclercq et al., 2014). Building on previous research, Riches et al. (2010) claimed that a SRT serves as an important tool in the diagnostic process of SLI. However, it is imperative to highlight that its validity as a potential clinical marker has not yet been evaluated systematically and fully.

While widely incorporated in language assessment tests (Dockrell and Marshall, 2015), the diagnostic accuracy of SRT has not been investigated for many languages, such as Greek,

including the Cypriot variety spoken in the Republic of Cyprus. Kamhi et al. (1984) already suggested that sentence repetition might produce more robust effects than spontaneous speech, and Everitt (2009) showed that it predicts later expressive abilities. This proposition followed the observation that children control their language productions by avoiding complex structures that are hard for them during spontaneous conversation. Consequently, in line with Seeff-Gabriel et al. (2010), we take it that a repetition task can be informative in terms of providing the full picture of children’s linguistic strengths and weaknesses.

During the last two decades, researchers have turned their interest to the diagnostic utility of the SRT and found that it is a good indicator of SLI, showing high levels of sensitivity and specificity for children speaking English (Conti-Ramsden et al., 2001), Cantonese (Stokes et al., 2006), French (Thordardottir et al., 2011; Leclercq et al., 2014), and dialects of English (Oetting et al., 2016). For example, Conti-Ramsden et al. (2001) investigated whether sentence repetition—along with a third person singular task, tense marking, and non-word repetition—could be a clinical marker for the identification of SLI in English. They found that the strongest marker among those examined was sentence repetition, with sensitivity and specificity values for sentence repetition at 90 and 85%, respectively.

A similar result was revealed by Stokes et al. (2006), who examined Cantonese-speaking children. Specifically, they found that sentence repetition can accurately differentiate children with SLI from their typically developing peers. Moreover, significant differences between a group of 20 children identified with SLI (aged 7.2–13.0) and two groups of typically developing children (chronologically matched and language-matched) were found by Briscoe et al. (2001). Furthermore, Botting and Conti-Ramsden (2003) investigated four groups of language-impaired children, including children with SLI, and concluded that sentence repetition discriminates children with SLI from the other groups, including typically developing children, better than non-word repetition and past tense tasks do.

Thordardottir et al. (2011) examined the accuracy levels in SLI identification for 5-year-old French-speaking children and showed that the SRT used was sensitive (86%) and specific (92%). Similarly, the accuracy of a SRT used by speech-language therapists for SLI identification in French was examined (Leclercq et al., 2014) and yielded high accuracy levels were yielded. In particular, the study showed that 97.1% of children with SLI and 88.2% of typically developing children were classified correctly. Riches et al. (2010) extended the populations under investigation in their study and examined three groups: a group of 14 adolescents with SLI (mean age: 15.3), a group of 16 autistic children who exhibited language impairment (mean age: 14.8), and a group of 17 typically developing adolescents (mean age: 14.4). The research demonstrated that sentence repetition serves as a sensitive marker for language impairments in both clinical populations, adolescents with SLI and autism spectrum disorder.

The importance of meaningful diagnostic accuracy levels is discussed by Komeili and Marshall (2013) who support that tests with high specificity and sensitivity can minimize misdiagnosis, in terms of both under- and over-diagnosis. A further issue

comes to light concerning the discrimination power of the task regarding age. Children between 3–6 and 6–11 years of age were tested on a repetition task and the results suggested that the younger children with SLI can be accurately identified in contrast to older children (Vender et al., 1981). Those findings were confirmed by research indicating that sentence repetition could be a sensitive clinical marker for younger children whose language abilities are incomplete, rather than for older children (Devescovi and Caselli, 2007). In contrast, the inclusion of complex sentences in a repetition task by Riches et al. (2010) showed that language-impaired individuals are identified even when they are adolescents. Other salient outcomes are those of Poll et al. (2010), who showed that sentence repetition is a good clinical marker of SLI in young adults.

Additionally the *type* of sentences included in a SRT has generated much discussion in the literature. Bernstein Ratner (2000) early on suggested that “[s]entences constructed at a level slightly above that observed in the child’s spontaneous speech are regularized in ways that reflect both the child’s extraction of form and meaning and the child’s linguistic capacity” (p. 293). She presupposes that for the construction of a task, researchers need to take into account not only the age of the children under investigation *per se*, but their language development stage as well. However, this is not always possible because for a considerable number of languages, no clear developmental trajectories are available regarding how children acquire sentence structures—and this includes Greek generally, and in particular the variety of interest in the current study, Cypriot Greek.

For the purposes of this study, complex morphosyntactic structures were selected for investigation under the assumption that children have already acquired simple structures. When sentences are long enough, the participant cannot simply copy them. As a result, they resort to the grammatical system in order to be able to repeat the sentences by processing, analyzing, and reconstructing their meaning. This can only happen if the participant has already acquired the grammatical structures (Marinis and Armon-Lotem, 2015), hence relatively long and complex sentences are used in a SRT. In other words, in order to repeat a sentence, a child has to know its syntax. Polišenská et al. (2015) confirmed that performance on sentence repetition depends on language ability and in particular, in the areas of morphosyntax and lexical phonology. However, a child will not repeat a sentence if it is not fully understood either (Vinther, 2002). Therefore, the grammatical structure needs to be acquired first in order to be comprehended and expressed.

The findings regarding the use of complex syntactic structures in SRTs are not surprising given the well-documented difficulties in using those structures in SLI (e.g., Leonard, 1998; van der Lely and Battell, 2003; Novogrodsky and Friedmann, 2006). Indeed, there are syntactic structures that are not easy to elicit (Seeff-Gabriel et al., 2010), such as question structures and passives, and consequently they have not yet been evaluated. Despite the known utility of the tasks regarding the elicited data, SRTs that include these structures have been subject to scant investigation (Riches et al., 2010).

## Some Background on Cypriot Greek

The Greek-speaking Republic of Cyprus, as it is summarized in Theodorou and Grohmann (2015), is generally described as “diglossia” (reviewed in Rowe and Grohmann, 2013), where the sociolinguistically “high” variety is typically accepted to be Standard Modern Greek (SMG), whereas the “low” variety is the vernacular Cypriot Greek (CG), of which Greek Cypriot is a native speaker. As can be accepted, the differences between the two varieties go far beyond the obvious aspects of language such as vocabulary, pronunciation, and prosody. Distinct differences between CG and SMG are lexical, phonetic, and (morpho)phonological properties of the language (a host of research since the seminal study of Newton, 1972). With regard to the morphosyntactic level are among others personal pronominal clitics, which precede the finite verb in SMG while CG employs enclisis in indicative declarative clauses (much work since Agouraki, 1997). For recent research on the syntax of CG-speaking children’s (a)typical language development, see among others Theodorou and Grohmann (2012) on relative clauses and Grohmann (2014a) for a review on clitics.

Because of the complex linguistic situation in Cyprus, the language status of Greek Cypriot children in this study is referred to as “bilinguals,” as by adopted Rowe and Grohmann (2013), a term that has been used by various other researchers in recent research on language acquisition and subsequent development (e.g., Kambanaros et al., 2013; Grohmann, 2014b; Antoniou et al., 2016; Theodorou et al., 2016; Grohmann et al., 2017). In this context, bilingualism is used to characterize the linguistic situation in Greek-speaking Cyprus: Children of Greek Cypriot parents, with CG-speaking family and friends, grow up with CG from birth and yet, are exposed to SMG from an early age. This usually comes first through children’s programme on TV, for example, and later through formal language instruction and interaction in public schools in all levels in SMG (though not necessarily in reality, as shown in Sophocleous, 2011; see also Leivada et al., 2017), thus enforcing exposure to SMG in a systematic way. Consequently, we further believe that language development in a bilingual context differs from very early on (Taxitari et al., 2015, 2017), both from monolinguals and bilinguals (Antoniou et al., 2016; Grohmann and Kambanaros, 2016).

The identification of language-impaired children in bilingual settings is not straightforward, since there are no screening or assessment tools specifically designed to diagnose impaired language in children who are CG-speakers (Kambanaros and Grohmann, 2013; Theodorou et al., 2016). Speech and language therapists (SLTs) as well as researchers usually rely on informal assessment measures, spontaneous language sampling, and clinical judgment to support the diagnostic process when formal diagnostic practices are not in place, a common phenomenon across a large number of EU countries (see Thordardottir, 2015). The diagnostic procedure becomes difficult not only because of the absence of appropriate screening and diagnostic tools for CG, it also creates confusion among policy-makers, teachers, and clinicians who may conceptualize both the language impairment itself and the need for speech and language services differently (Kambanaros and Grohmann, 2013).

In a more recent study (see also Theodorou, 2013; Theodorou et al., 2013), Theodorou et al. (2016) examined a number of norm-referenced tests published for SMG that assess the language abilities of monolingual children in Greece. These tests were modified into CG to address dialectal differences. The full assessment battery included measures of receptive vocabulary, comprehension and production of morphosyntax, metalinguistic concepts, sentence repetition, narrative retelling, articulation and phonological processing, word definitions, sound distinctions, and word finding. The study suggests that a combination of existing diagnostic tools support the diagnostic procedure when modified for CG on the basis of acceptable accuracy levels. This in turn allows the assumption that, if clinicians adopt the combinations suggested in that study, the likelihood for a correct diagnosis increases. The importance of accurate detection reflects on appropriate intervention, which has been acknowledged by several researchers (Fey and Cleave, 2008; Gallagher and Chiat, 2009).

This study addresses the question whether a SRT that elicits complex syntactic structures can serve as an accurate screening task for the identification of children who need further language assessment. Secondly, it will be evaluated whether there are qualitative differences in terms of morphosyntactic errors produced by children.

## METHODS AND MATERIALS

### Participants

Participants were 38 CG-speaking children aged 5–9 years who completed a SRT as part of a larger study about diagnosis of SLI in CG (e.g., Theodorou and Grohmann, 2015; Theodorou et al., 2016). The children were divided into four groups. Nine children were included in the younger group of children with SLI (SLI-Y: 7 boys and 2 girls, mean age 5.6, *SD* 0.3), and seven in the older group (SLI-O: 3 boys and 4 girls, mean age 7.8, *SD* 0.8). Ten participants were included in the younger group of TLD children (TLD-Y: 6 boys and 4 girls, mean age 5.8, *SD* 0.6) and twelve in the older group (TLD-O: 6 boys and 6 girls, mean age 7.10, *SD* 0.6). Building on our previous work (Theodorou et al., 2016), we compare the two groups of children with SLI to chronological age-matched groups following the proposed practice in assessing the accuracy of clinical markers (Plante and Vance, 1994; Bortolini et al., 2002, 2006). The background information on the 38 participating children is reported in **Table 1**.

**TABLE 1** | Participant details.

Group	Age range	Number of participants	Mean	Stand. dev.	Gender
TLD-Y	4.5–6.6	10	5.8	0.6	6M, 4F
TLD-O	6.7–8.7	12	7.10	0.6	6M, 6F
SLI-Y	4.11–5.11	9	5.6	0.3	7M, 2F
SLI-O	6.7–8.1	7	7.8	0.8	3M, 4F

TLD, children with typical language development; SLI, children with specific language impairment; Y, younger; O, older.

Subject selection criteria included: (i) CG-speaking background, (ii) no history of neurological, emotional, developmental, or behavioral problems, (iii) hearing and vision adequate for test purposes, (iv) performance within a broad range of normal on a measure of non-verbal intelligence (Raven's Coloured Progressive Matrices, Sideridis et al., 2015), and (v) no gross motor difficulties. All information was obtained either from speech therapists and teachers or from their parents. The children came from families with a medium to high socioeconomic status as measured by mother's education level using the European Social Survey (2010) database. Background information on the participating children is reported in **Table 2**.

Adopting the notion of "(discrete) bilectalism" from Rowe and Grohmann (2013), we consider "monolingual" children in diglossic speaker communities to be (at least) bilectal in the "high" and "low" varieties (see Kambanaros et al., 2013 for the first published study on child language implementing this term). With respect to the children participating in the present study, however, we can confidently state that they were all bilectal in CG (the native variety, spoken at home) and SMG (introduced formally in preschool; language of media and communication)—as understood through the works just cited. In particular, no children were simultaneous or sequential acquirers of an additional language and no child was a native speaker of SMG or received, to the best of our knowledge, any more input of strict SMG than any other.

**Table 3** illustrates the performance of the children on the Raven's Progressive Matrices test (non-verbal IQ test) (Raven et al., 1998; Sideridis et al., 2015). Subject selection criteria included normal performance on the non-verbal IQ test. This requirement is satisfied for each child separately and there are no statistically significant differences in non-verbal IQ between the SLI groups and the controls.

Children with SLI were recruited through private speech therapy clinics based on a protocol that included the previous identification of the participants by certified SLTs based on case history information, informal testing of comprehension and production, analysis of spontaneous language samples, and clinical observation. The diagnosis was later confirmed by a battery of tests developed for the assessment of SLI in Cyprus (Theodorou et al., 2016). The full assessment battery included measures of receptive vocabulary, comprehension and production of morphosyntax, metalinguistic concepts, sentence repetition, narrative retelling, articulation and phonological processing, word definitions, sound distinctions, and word finding. The groups' results on those tests are tabulated in Appendix A in Supplementary Material. The reader can find a detailed description of the recruitment procedure and complete descriptions of the tests in Theodorou et al. (2016).

### Sentence Repetition Task (SRT)

The ability of children to repeat syntactically complex sentences was assessed with an SRT, thus adopting the suggestion (Redmond, 2005; Stokes et al., 2006) that the stimuli of such a task should be complex in order to avoid ceiling performance. Accordingly, complex structures that are used frequently in CG, as in SMG were chose for inclusion. Indeed, it is important

TABLE 2 | Participants' details.

Group	Age range	No. of participants	Mean (SD)	Sig. (2-tailed)-Age	Gender	Mo's ed. (SD)	Sig. (2-tailed)-Mo's ed.
TLD	4.5–8.7	22	6.10 (1.3)	0.29	12M, 10F	3.95 (1.1)	0.06
SLI	4.11–8.1	16	6.2 (1.3)		10M, 6F	3.37 (0.69)	

\*The mean difference is significant at the 0.05 level. SD, standard deviation; TLD, children with typical language development; SLI, children with specific language impairment; Y, younger; O, older; M, male; F, female; Mo's ed., mother's education (0 = did not complete primary education, 1 = completed primary education, 2 = completed high school, 3 = completed lyceum, 4 = diploma, 5 = university degree, 6 = master qualifications, 7 = PhD qualification).

TABLE 3 | Means, standard deviations, and significant levels of all groups (Raven's).

Groups	Mean scores (standard deviation)		Sig. (2-tailed)
	TLD (n = 10)	SLI (n = 9)	
Younger	90 (12.47)	100.56 (12.86)	0.087
	TLD (n = 12)	SLI (n = 7)	
Older	94.58 (9.64)	95.71 (17.66)	0.880

TLD, children with typical language development; SLI, children with specific language impairment.

to note that for task construction and grading of structural difficulty, no model was adopted, because there is no relevant literature either for CG or for SMG. However, the items included represent structures that can be produced by typically developing children that are SMG speakers, as shown in corpora studies. Summing up, Mastropavlou and Tsimpli (2011) conclude that *na*-clauses can be produced even at the age of 2. Emergence of *pu*-relatives and *oti*-clauses follow later. Further, the structures included are those that have been found to be problematic for children with SLI either in Greek (including CG) (Stavarakaki, 2001; Theodorou and Grohmann, 2012) or in other languages, as the international literature (e.g., Leonard, 2001; Friedmann and Novogrodsky, 2004; Kunnari et al., 2014) suggests. The test consists of 24 items exploring the imitation of structures within six syntactic categories with four examples of each type: object relative clauses (1), subject relative clauses (2), embedded *oti* "that"-clauses (3), adjunct *giati* "because"-clauses (4), negative *den*-sentences (5), and subjunctive *na*-clauses (6).

- (1) Vlepo ti <sup>n</sup>gota pu a<sup>n</sup>gałazi i çata.  
*I am watching the hen that the cat is hugging.*
- (2) Akouis to maθiti pu lali tin istoria.  
*You are hearing the pupil who is telling the story.*
- (3) Ipes oti i jaja emairepse su to fai.  
*You said that granny cooked your food.*
- (4) I daskala tu eçirokrotise <sup>n</sup>don jati itan θcevazmenos.  
*His teacher applauded him because he was studious.*
- (5) O mixalis e <sup>n</sup>do epline to proi.  
*Michalis didn't wash it in the morning.*
- (6) Prepi na mu to ðocis sto parko.  
*You must give it to me at the park.*

Specific language properties of CG were taken into consideration for the test design, including syntactic (e.g., clitics appear

post-verbally: *eçirokrotise<sup>n</sup>don* in CG, *ton çirokrotise* in SMG), phonological (e.g., consonant deletion: *emairepse* in CG, *majirepse* in SMG), and morphological aspects (e.g., syllabic augment [e] in past tense: *eçirokrotise* in CG, *çirokrotise* in SMG), among others (see Appendix B in Supplementary Material). The length of the sentences was between 9 and 13 syllables (mean: 15.54, SD: 4.34), which resembles sentences appearing in fairy-tales for pre-primary school level as well as the length of sentences appearing in text books grade 1. As for the vocabulary used, every day words and words that are frequently used in fairy tales and in the text books of grade 1 were selected, to avoid the vocabulary content having an undue influence on the sentence repetition ability (Polišenská et al., 2015). In particular, nouns and verbs were restricted to early-acquired words, such as "mum," "granny," "baby," "food," "want," "say," and "wash."

## Procedure

The participants were asked to listen to 24 pre-recorded sentences. After each sentence, they were asked to repeat it as close to the original as possible. The stimuli were audio-recorded to ensure that all participants heard the sentences in the same way and presented via a PC in a fixed order using Power Point. The children were tested individually by trained research assistants. The examiner sat at a table either next to or opposite the children and said: "You are going to hear a sentence while you are watching the computer screen. You have to say exactly what you have heard." On the computer screen a green circle would appear in order to keep the attention of the child away from other distractions in the room. No feedback was provided during the actual experiment, but encouragement was given when deemed necessary. Children's responses during the administration of the experimental task were audio-recorded using an Olympus WS-311M digital voice-recorder with a high-quality built-in microphone. These recordings were used to transcribe the children's responses for subsequent scoring.

## Scoring

Two different methods of scoring were examined. This decision was driven by Redmond's (2005) claim that in order for a task to be included in a battery aiming to detect SLI, a more refined scoring procedure is required. Consequently, the responses first were scored as correct (1 point) when a sentence was repeated exactly, with all the sentence elements included (hereafter Scoring Method 1). Scoring Method 1 mirrors that used for the TOLD-P3 Sentence Imitation subtest (Newcomer and Hammill, 1997) as well as the method adopted by Stokes et al. (2006) and Rispens (2004). Hence, the possible score

range using this method was 0–24. For the second scoring method (hereafter Scoring Method 2), responses were scored according to the number of errors made in each sentence in agreement with the system developed for CELF-R (Semel et al., 1989), which was also used by Conti-Ramsden et al. (2001). That is, items were scored on a 0–3 scale, with 3 representing an exact repetition, 2 a sentence repetition with 1 error, 1 with 2 or 3 errors, and 0 with more than three errors. The maximum possible score using Scoring Method 2 was thus 72. For both scoring methods, phonological errors were not taken into consideration since the vast majority of the children with SLI exhibited some phonological difficulties as their performances for the phonological test indicate (see Appendix A in Supplementary Material). At this point, it is important to clarify that phonological processes used by our participants did not interact with calculated errors. For example, a common phonological process used was syllable deletion in multisyllabic words (e.g., [epakoluθusan] instead of /eparakoluθusan/ “they were watching”).

## Error Analysis

In order to get some qualitative insights with regards to the morphosyntactic errors made by the participants a broad error analysis was followed. That is, each of the sentences produced was classified as syntactically correct either identical to the prompt or not. Then the errors or alternatives provided were classified as omission (7), substitution (8), addition (9), and change of word order (10) (Note that if the substitution resulted due to a phonological process only, it was not considered an error). A more detailed analysis followed to determine the affected linguistic element. Specifically, whether the error concerned a content word (7), free-standing morpheme (8), or an inflectional grammatical morpheme (11).

*Target sentence:* Vlepo tin <sup>ɹ</sup>gota pu a<sup>n</sup>galiazi i γata.  
“I am watching the hen that the cat is hugging.”

*Produced sentence:*

- (7) Vlepo tin (omission/content word) pu a<sup>n</sup>galiazi i γata.
- (8) Vlepo tin <sup>ɹ</sup>gota **na** (substitution/free-standing morpheme) a<sup>n</sup>galiazi i γata.
- (9) Vlepo tin <sup>ɹ</sup>gota pu **tin** (addition/free-standing morpheme) a<sup>n</sup>galiazi i γata.
- (10) Tin <sup>ɹ</sup>gota **vlepo** (change of the word order) pu a<sup>n</sup>galiazi i γata.
- (11) Vlepo tin <sup>ɹ</sup>gota pu a<sup>n</sup>galiaze (substitution/free-standing morpheme) i γata.

## RESULTS

### Group Differences

The performance of the four groups was compared according to the two scoring methods, provided in **Table 4**.

The differences on performance between children with SLI and TLD peers, with SLI scoring lower than TLD for both scoring methods, is graphically depicted in **Figure 1** (Scoring Method 1) and **Figure 2** (Scoring Method 2). To examine whether the task yielded significant differences between the groups, a one-way

**TABLE 4** | Group performances on the SRT.

Scoring method	Group	Mean	SD
1 (out of 24)	TLD-Y	14.6	3.098
	TLD-O	18.2	4.366
	SLI-Y	7.9	3.790
	SLI-O	11.0	5.164
2 (out of 72)	TLD-Y	57.6	5.777
	TLD-O	63.5	7.379
	SLI-Y	40.2	13.890
	SLI-O	49.9	9.668

TLD, children with typical language development; SLI, children with specific language impairment; Y, younger; O, older.

ANOVA was conducted. The test revealed significant differences between the groups for both methods, Scoring Method 1 [ $F_{(3, 34)} = 11.92, p = 0.00$ ] and Scoring Method 2 [ $F_{(3, 34)} = 11.47, p = 0.00$ ].

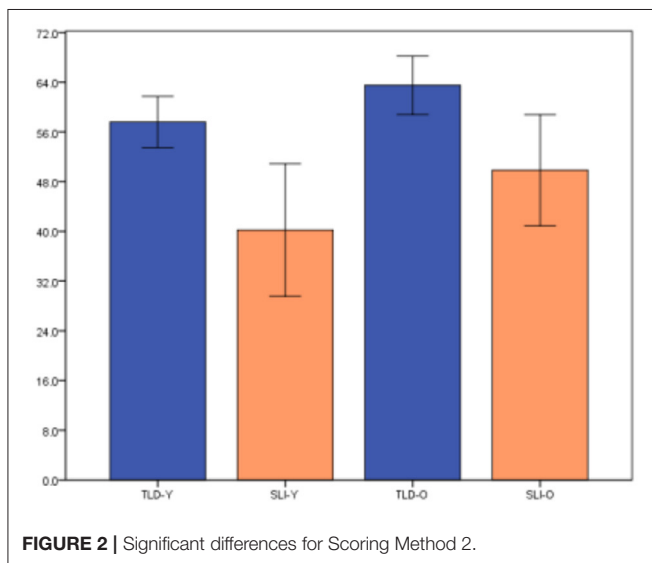
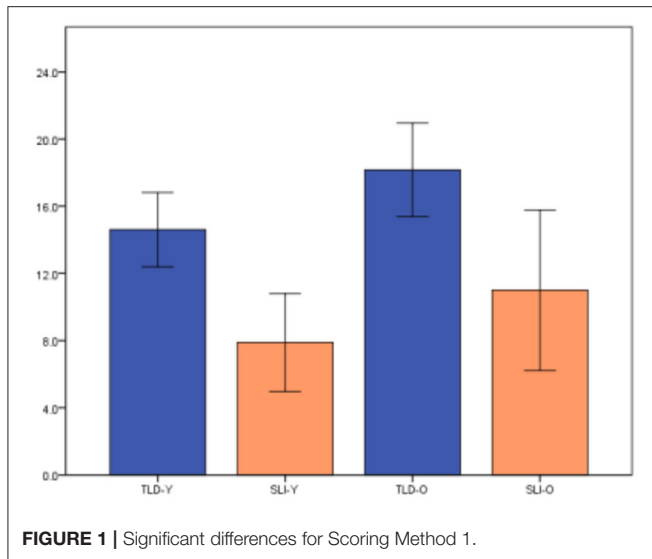
A two-way ANOVA was conducted to examine the effects of age (Old vs. Young) and language group (TLD vs. SLI) on the two scoring methods. For the first scoring method, both the main effect of age [ $F_{(1, 34)} = 6.072, p = 0.019$ ] and the main effect of language group [ $F_{(1, 34)} = 26.226, p < 0.001$ ] were significant. These results indicate that the TLD participants ( $M = 6.10, SD = 1.3$ ) performed significantly higher than the SLI participants ( $M = 6.2, SD = 1.3$ ). A non-significant interaction [ $F_{(1, 34)} = 0.028, p = 0.867$ ] implies that the effect of language group was the same across the old and young participants.

Similar results apply for the second scoring method. Both the main effect of age [ $F_{(1, 34)} = 6.247, p = 0.017$ ] and the main effect of language group [ $F_{(1, 34)} = 24.907, p < 0.001$ ] were significant and their corresponding interaction was not significant [ $F_{(1, 34)} = 0.361, p = 0.552$ ]. Again, the TLD participants ( $M = 6.2, SD = 1.3$ ) performed significantly better than the SLI participants ( $M = 6.10, SD = 1.3$ ) and the effect of language group was the same across the old and young participants. Interactions for scoring method 1 and scoring method 2 are illustrated in **Figures 3, 4**, respectively.

Summarizing so far, in line with other studies, CG-speaking children with SLI performed significantly below the TLD groups, rendering the SRT a potential clinical marker. Interestingly, the children’s performance did not differ as a function of age, thus permitting the treatment of the participants as two groups, children with SLI and TLD children, for the remainder of the analysis.

### Specificity and Sensitivity

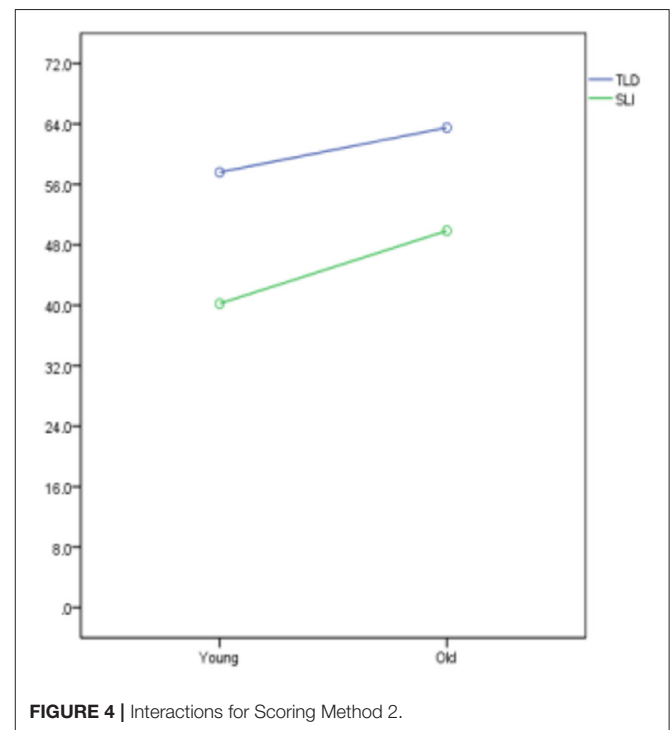
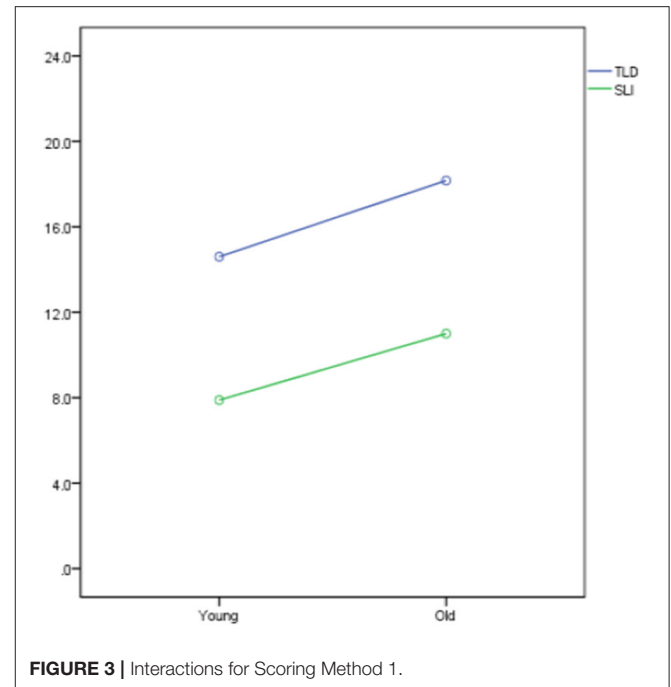
It is already known that the significant differences between the groups are not reliable enough to characterize the SRT as an accurate tool for the detection of the impairment (Plante and Vance, 1994). Consequently, we proceeded to evaluate the sensitivity and specificity of the task used by conducting binary logistic regression analysis. More specifically, the analysis was carried out in order to show whether the children can be classified as children with SLI or TLD children, according to their



performance in this task, for either of the two scoring methods or a combination of the two.

The results of the logistic regression analyses are tabulated in **Table 5**, where the percentages and the number of children that were correctly classified are shown for all three scoring arrangements.

Scoring Method 1 seems to be more accurate than Scoring Method 2, whilst the combination of the two scoring methods reveals an identical accuracy level to Scoring Method 1. It appears that Scoring Method 1 can classify TLD children, as such, with 81.8% specificity, but it cannot classify SLI children equally well, as the reported sensitivity level is only 75%. Moreover, Scoring Method 1 can classify children with SLI at 78.9% accuracy. Summarizing so far, it is observed that Scoring Method 1 is an accurate discriminator for CG-speaking children with SLI, although the sensitivity level, in line with Plante and Vance (1994), cannot be characterized as adequate.



However, there is an issue that needs to be taken into consideration. One child belonging to the group of older children with SLI scored very high on this task, in contrast to his low performance in the other tasks, included in the diagnostic battery. This participant was a boy of 8.6 years who scored 22 out of 24 for Scoring Method 1 and 70 out of 72 for Scoring Method 2. His performance stands in stark contrast to the other children's



performance included in the group, given the fact that the child whose performance followed his scored 12 and 53 on the two methods, respectively. Given this observation, we treated this particular child as an outlier and ran the regression analysis once more excluding him. **Table 6** illustrates the percentages and the numbers of children that were correctly classified for each of the scoring methods as well for the combination of the methods as well, after the child was dropped from the analysis.

It is interesting to note that the accuracy levels shifted slightly upwards. **Table 6** shows that both scoring methods can classify accurately (81.1%) both groups, the children with SLI (sensitivity: 80%) and TLD children (specificity: 81.8%). However, with regards to the combination of the two methods, a slight reduction in the accuracy level is noted. A general outcome is that SRT can serve as a screening task for SLI identification. However, more research is needed, with more attention due to the design of the experiment.

## Morphosyntactic Structures

The performance of children with SLI and their TLD peers in terms of correct raw scores on sentence repetition according to grammatical structure are graphically depicted in **Figure 5** (individual results appear in Appendix C in Supplementary Material). It is observed that TLD children do not perform ceiling on the SR task. This is expected given that the stimulus included in the task are complex. Furthermore, and at least for research on relative clauses in CG (Theodorou and Grohmann, 2012), TLD children have not fully acquired them even at the age of 9 years old.

To examine whether significant differences yield between TLD children and children with SLI, *t*-tests were conducted. The analysis shows significant differences for the younger groups, between TLD-Y and SLI-Y, in object relative clauses [ $T_{(17)} = 2.918, p = 0.01$ ], subject relative clauses [ $T_{(17)} = 5.178, p = 0.00$ ], embedded *oti* “that”-clauses [ $T_{(17)} = 3.444,$

$p = 0.003$ ], negative *den*-sentences [ $T_{(17)} = 2.109, p = 0.05$ ], and subjunctive *na*-clauses [ $T_{(17)} = 3.820, p = 0.001$ ]. As for the older groups, significant differences were found between TLD-O and SLI-O in object relative clauses [ $T_{(17)} = 2.846, p = 0.011$ ], embedded *oti* “that”-clauses [ $T_{(17)} = 3.259, p = 0.005$ ], negative *den*-sentences [ $T_{(17)} = 2.342, p = 0.032$ ], and adjunct *giati* “because”-clauses [ $T_{(17)} = 2.712, p = 0.015$ ]. Analysis was carried out to examine whether significant differences were revealed between younger and older groups of children. A significant difference was detected between TLD-Y and TLD-O in terms of object relative clauses [ $T_{(20)} = -2.428, p = 0.025$ ]. As for the comparisons between SLI-Y and SLI-O, analysis showed that there are significant differences in subject relative clauses [ $T_{(14)} = -2.191, p = 0.046$ ] and subjunctive *na*-clauses [ $T_{(14)} = -2.138, p = 0.051$ ].

## Error Analysis

Acknowledging that sentence repetition allows for a collection of qualitative information about different language levels (Komeili and Marshall, 2013), for the purposes of the current study we investigate the errors made in terms of quantity. This is because of the main aim of the study, which is the evaluation of the SRT as a language-screening tool for CG-speaking children. Consequently, one of the scoring procedures followed by Stokes et al. (2006) was broadly applied, where the core elements of a sentence are isolated and then scored accordingly. First, the sentences produced were classified as syntactically correct or incorrect independently from the target sentences such as (12).

- (12) Target sentence: Akuis to mathiti pu lali tin istoria.  
 “You are listening to the pupil who is telling the story.”  
 Produced sentences: Akuis **ena** mathiti pu lali tin istoria.  
 “You are listening to a pupil who is telling the story.”

A one-way ANOVA was conducted which shows significant differences between the groups [ $F_{(3, 34)} = 9.682, p = 0.00$ ]. In order to find out whether there was a difference among the groups, a *post-hoc* Scheffé test was applied. The results show significant differences between younger children with SLI and younger TLD children ( $p = 0.004$ ), whereas the difference between older children with SLI and older TLD children is not significant ( $p = 0.073$ ).

Moving to a more detailed analysis, the errors made were classified as Omissions, Substitutions, Additions, and Word Order Error. As **Figure 6** illustrates, differentiation between groups can be observed. To examine whether errors made yielded significant differences between the groups, a one-way ANOVA was conducted. The test reveals significant differences for all four types of errors [Omissions:  $F_{(3, 34)} = 10.059, p = 0.00$ ; Substitutions:  $F_{(3, 34)} = 8.170, p = 0.00$ ; Additions:  $F_{(3, 34)} = 5.732, p = 0.003$ ; and Word Order Errors:  $F_{(3, 34)} = 3.864, p = 0.018$ ].

In order to discover the groups that differ significantly, a *post-hoc* Scheffé test was conducted. Regarding Omissions, a significant difference was yielded between younger children with SLI and younger TLD children ( $p = 0.004$ ) as well as between

**TABLE 5** | Percentages (and number of children) correctly classified by each scoring method.

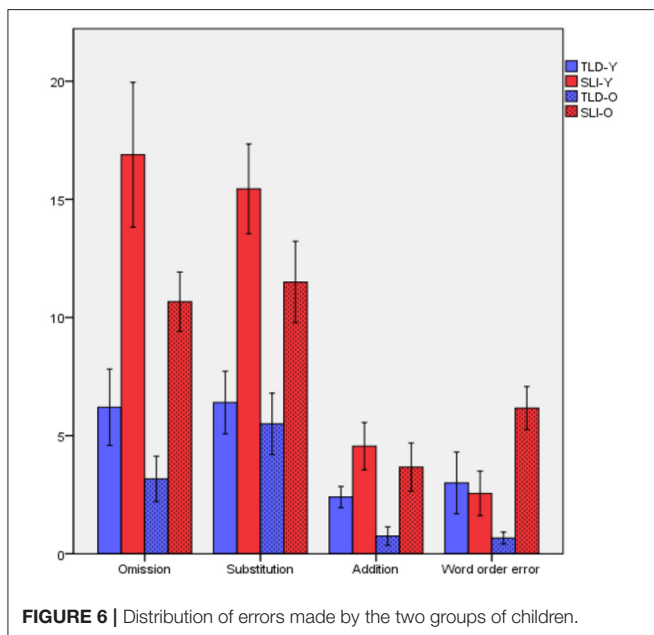
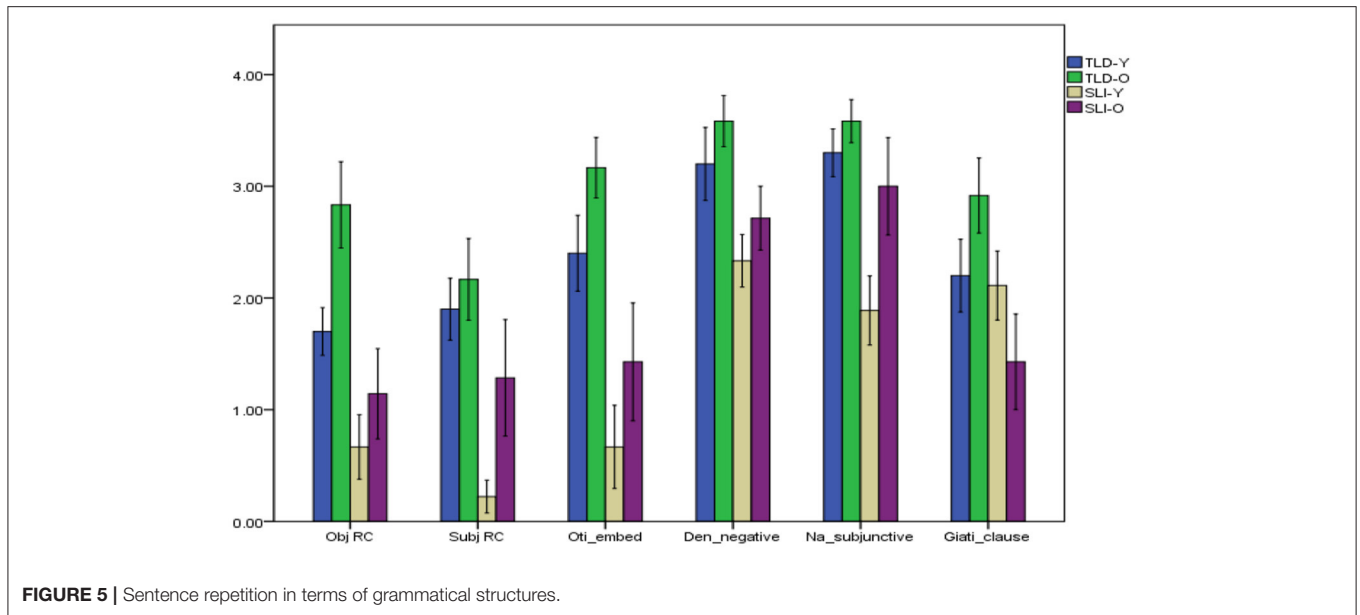
Scoring method	Children with SLI (sensitivity)	TLD children (specificity)	Overall accuracy
1	12/16 (75%)	18/22 (81.8%)*	30/38 (78.9%)
2	12/16 (75%)	17/22 (77.3%)	29/38 (76.3%)
1 + 2	12/16 (75%)	18/22 (81.8%)*	30/38 (78.9%)

\*\* Good discriminant level, \* Fair discriminant level.

**TABLE 6** | Revised percentages (and number of children) classified by each scoring method.

Scoring method	Children with SLI (Sensitivity)	TLD children (Specificity)	Overall accuracy
1	12/15 (80%)*	18/22 (81.8%)*	30/37 (81.1%)
2	12/15 (80%)*	18/22 (81.8%)*	30/37 (81.1%)
1 + 2	11/15 (73.3%)	18/22 (81.8%)*	29/37 (78.4%)

\*\* Good discriminant level, \* Fair discriminant level.



younger children with SLI and older TLD children ( $p = 0.000$ ). Significant differences are also observed between younger TLD children and younger SLI ( $p = 0.004$ ) and between younger SLI and older TLD ( $p = 0.001$ ) in terms of Substitutions. In relation to Additions, the analysis shows significant difference only between younger children with SLI and older TLD children ( $p = 0.003$ ). Moreover, older children with SLI differ significantly from older TLD children in terms of Word Order Errors ( $p = 0.02$ ). It is highlighted here that no significant difference is detected between younger and older children in both cases, i.e., children with SLI and TLD children do not differ within the age groups for any of the error types.

Going a step further, we examined which morphological elements are affected in the produced sentences. To this end, the affected element—content word, free-standing morpheme, inflectional morpheme—was determined for each error. **Table 7** presents the mean and standard deviation of the affected elements for each type of errors for all groups.

A one-way ANOVA was conducted to examine whether the affected elements are different for each group of participants. Significant differences were yielded between the groups for omission of content words [ $F_{(3, 34)} = 7.444, p = 0.001$ ], omission of free-standing morphemes [ $F_{(3, 34)} = 10.515, p = 0.00$ ], substitution of content words [ $F_{(3, 34)} = 6.117, p = 0.002$ ], substitution of inflectional morphemes [ $F_{(3, 34)} = 7.902, p = 0.00$ ], addition of content words [ $F_{(3, 34)} = 3.612, p = 0.023$ ], addition of free-standing morphemes [ $F_{(3, 34)} = 4.326, p = 0.011$ ], and change in the order of free-standing morphemes [ $F_{(3, 34)} = 5.375, p = 0.004$ ]. The analysis continued with determining the pair of groups that differ significantly in terms of the affected morphological elements. They were found to differ significantly when a *post-hoc* Scheffé test was conducted. The results are provided in **Table 8**.

## DISCUSSION

Research efforts on children with SLI have suggested sentence repetition capabilities can be a clinical marker. The primary interest regarding this study was to investigate whether SRT could serve as a screening task for bilingual CG-speaking children with SLI. The second aim was to identify the relation between SRT and a group of valid language tests included in a language assessment battery recently examined by the authors (Theodorou et al., 2016). Further analysis followed to examine the differences in terms of morphosyntactic errors produced by the participants.

Summing up, the SRT yielded significant differences in performance of CG-speaking children with SLI and those with

TABLE 7 | Mean (standard deviation) of affected elements.

	Omissions			Substitutions			Additions			Word order errors		
	Content word	Free-standing morpheme	Inflectional morpheme	Content word	Free-standing morpheme	Inflectional morpheme	Content word	Free-standing morpheme	Inflectional morpheme	Content word	Free-standing morpheme	Inflectional morpheme
TLD-Y	2.1 (2.2)	4.2 (3.2)	0 (0)	2.4 (1.7)	3 (3.1)	1 (0.9)	0.2 (0.4)	2.2 (1.3)	0 (0)	1.2 (1.7)	1.8 (2.6)	0 (0)
SLI-Y	6.2 (4.0)	10.3 (5.1)	0.3 (1.0)	5.9 (3.1)	4.3 (1.9)	5.2 (3.8)	1.1 (1.2)	3.2 (2.3)	0.2 (0.4)	1.0 (1.5)	1.6 (1.5)	0.0 (0.0)
TLD-O	0.9 (1.2)	2.1 (1.9)	0.2 (0.6)	2.1 (1.8)	2.4 (2.1)	1.0 (1.3)	0.1 (0.3)	0.7 (1.4)	0.0 (0.0)	0.1 (0.3)	0.6 (0.8)	0.0 (0.0)
SLI-O	3.2 (2.8)	7.5 (2.1)	0.0 (0.0)	4.5 (1.2)	5.0 (2.4)	2.0 (2.0)	0.8 (1.2)	2.8 (1.5)	0.0 (0.0)	1.3 (1.0)	4.8 (1.6)	0.0 (0.0)

TABLE 8 | Pairs of groups that differ significantly in terms of types of errors.

	Pairs	Sign. level
Omission of content word(s)	TLD-Y/SLI-Y	$p = 0.017$
	TLD-O/SLI-O	$p = 0.001$
Omission of free-standing morphemes	TLD-Y/SLI-Y	$p = 0.006$
	TLD-O/SLI-Y	$p = 0.000$
Substitution of content word(s)	TLD-Y/SLI-Y	$p = 0.015$
	TLD-O/SLI-Y	$p = 0.005$
Substitution of inflectional morphemes	TLD-Y/SLI-Y	$p = 0.003$
	TLD-O/SLI-Y	$p = 0.002$
Addition of content word(s)	SLI-Y/SLI-O	$p = 0.0041$
	SLI-Y/TLD-O	$p = 0.046$
Addition of free-standing morphemes	TLD-O/SLI-Y	$p = 0.016$
Word order error: Free-standing morphemes	TLD-O/SLI-O	$p = 0.004$

TLD. The outcome confirms previous research findings for other languages, such as English (Conti-Ramsden et al., 2001; Seeff-Gabriel et al., 2010; Redmond et al., 2011), Cantonese (Stokes et al., 2006), Italian (Devescovi and Caselli, 2007), and French (Thordardottir et al., 2011; Leclercq et al., 2014), thus revealing that sentence repetition could be an effective clinical marker for bilectal CG-speaking children. We wish to highlight that the SRT used factored in dialectal (or variety) issues (Oetting et al., 2016) in the context of diglossia. Moreover, the majority of the grammatical structures used in the task was found to differentiate the performance of TLD children from their peers with SLI. This study is the first research to investigate sentence repetition in CG and therefore, further research is needed for a more complete picture.

The group differences found motivated the evaluation of the discrimination accuracy of the task. The high sensitivity and specificity levels which have been found for other languages, for example, English (Conti-Ramsden et al., 2001), are not replicated here, which may be due to the task design among other reasons that are discussed below. However, nearly accurate enough levels for Scoring Method 1 have been yielded (and slightly lower levels for Scoring Method 2).

Given the fact that sentence repetition has been found to be related to measures examining grammatical skills, namely, phonology, morphosyntax and semantics, an error analysis was conducted to compare the morphosyntactic abilities of the participants. Our findings allow us to directly support the claim put forward in the relevant literature (Lust et al., 1996; Marinis and Armon-Lotem, 2015; Polišenská et al., 2015) that the performance on sentence repetition is an indicator of a child's grammatical ability.

Other notable observations touch upon the errors made in terms of affected morphological errors-content words, free standing morphemes, inflectional morphemes. As for content words, though found to be affected, the differences between the groups are marginal, whereas more significant differences are observed for both free-standing and inflectional morphemes between the groups. Interestingly, no omission of inflectional morphemes was found which is arguably owed to the morphological richness of the Greek language where each lemma is usually highly inflected.

Another interesting revelation from the error analysis concerns the strategy of the older children with SLI (SLI-O) to produce alternative grammatically correct structures instead of the exact wording of what was heard. We can thus conclude that bilingual CG-speaking children with SLI do not produce ungrammatical sentences, but rather resort to structures that are accessible to them—even when considerably complex.

Summing up so far, the tool presented here could be adopted by SLTs as a screening task for identifying children who need further language assessment accurately. It is possible also for early education specialists (e.g., teachers) to be trained on the use and interpretation of the tool. This, in turn, would facilitate access to the appropriate services for language-impaired children. A short identification task would minimize the risk of non-identification and inaccessibility appropriate intervention, as has previously been recommended regarding evaluation protocols (Redmond et al., 2011).

The outcome of the task permits us to make a suggestion about the distinction of the discrimination power of the task in relation to the age of the children, in that younger children with SLI are differentiated more accurately than older ones (Vender et al., 1981; Devescovi and Caselli, 2007) has not been confirmed here. What is relevant is that older children with SLI produced syntactically correct sentences not identical to what they heard. The findings here tend to corroborate the suggestion by Riches et al. (2010) that SRT can identify older language-impaired children. It is assumed that the diagnostic accuracy has to do more with the type of the structures included in the task, rather than the task as such and is in agreement with Leclercq et al. (2014), who contend that SRT is very complex for children with SLI.

Apart from the matter of identification, some theoretical issues could also be addressed. Besides carrying out an analysis for both groups of TLD and language-impaired children, further analysis comparing younger and older groups did not reveal any significant difference. This outcome suggests that, at least for the set of structures included here, age does not play a role given that only minimal developmental progress is reported for children with SLI and for TLD children. Whilst the finding needs to be interpreted with caution, we contend that Greek Cypriot children, even at the age of 9, are still developing their language skills. As a consequence of this observation, we have insufficient evidence to make a definitive contribution to the ongoing debate pertaining to delay vs. deviance.

Additionally, researchers have highlighted several advantages of the task. First, it is claimed that SRT can be easily administered and analyzed (Lust et al., 1996), allowing for the evaluation of specific grammatical structures under controlled situations. That is, given the fact that it is implemented using a one-to-one format, this provides the opportunity for examiners to control the conditions in which children complete the task. In addition, a structured repetition task allows the investigator to select the target sentences carefully, according to the specific aims of the research, whereas this is not always possible if a spontaneous speech sample is evaluated. Thus, the researcher can examine morphosyntactic structures that are not easy to elicit either in spontaneous language or in other structured elicitation tasks. In addition, it is a natural skill that needs little effort and even young

children recall sentences willingly. Moreover, it is postulated that the task does not seem to be influenced by factors, such as gender (Seeff-Gabriel et al., 2010). Concerning the relation between socioeconomic status and sentence repetition ability the existing evidence is contradictory, since there are studies that have contended there is a relation between high SES and better performance on SRT (Roy et al., 2014; Balladares et al., 2016), whilst others have reported no such influence (Gardner et al., 2006).

Some limitations of this investigation are reported as follows. First, the sample size is small and the age range quite large. However, sample size seems to be in line with the relevant published literature, such as Stokes et al.'s (2006) 16 and Seeff-Gabriel et al.'s (2010) 13 children with SLI investigated. Second, an issue that came to light concerns the construction of the task. We now believe that in the future, a replication of a tool to examine sentence repetition ability should take into consideration issues about language development and language impairment in CG (and SMG), such as structures that are expected to be developed by the ages under examination, rather than only the complexity parameter. By so doing, the task will become even more specific to structures that are documented as being problematic in the present study and previous research for CG (Theodorou, 2013; Theodorou et al., 2016). In addition, in order for the task to be administered for screening purposes, cut-off points should be established (Stokes et al., 2006), based on previous research Conti-Ramsden et al. (2001). Unfortunately, so far no standardized tests have been established for CG, although a battery of tests were found to be accurate in the diagnosis of SLI (Theodorou et al., 2016).

Another research direction could be the evaluation of SRT for measuring the progress of language intervention programs (Devescovi and Caselli, 2007). If there is evidence-based research that the SRT can really measure therapy progress, then the benefits will be two-fold. First, it could be a tool for SLTs to measure the effectiveness of the intervention. Second, policy-makers would then have tangible data to support the need for speech-language therapy services for those children with language difficulties. It is imperative to point out that the SRT presented here is not available to speech-language therapists yet, but a revised version could be in the future.

## CONCLUSION

It is crucial for clinicians and researchers alike to be sufficiently confident about the identification accuracy of a task used to identify children who experience SLI. However, no language test is able on its own to diagnose and describe the language abilities of a child in full and of course, none is sufficient to formulate recommendations for therapeutic intervention (Dockrell, 2001). Research has shown that sentence repetition is a useful tool for identifying children's language skills alongside other language tests. This study aimed to shed some light on the question whether children with SLI can be identified by using an SRT in the context of diglossia in Cyprus, where no diagnostic tests designed for the particular situation are available, and the results suggest such a task could be a potential clinical marker for SLI in CG. The outcome of this study is

indicative and can be considered as a starting point for additional research.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Center of the Educational Research and Assessment of Pedagogical Institute of Cyprus with written informed consent from all subjects. All subjects gave written informed consent. The protocol was approved by the Center of the Educational Research and Assessment.

## REFERENCES

- Agouraki, Y. (1997). "On the enclisis/proclisis alternation," in *Proceedings of the 2nd International Conference on Greek Linguistics*, eds G. Drachman, A. Malikouti Drachman, J. Fykias, and C. Klidi (Graz: W. Neugebauer Verlag), 393–404.
- Antoniou, K., Grohmann, K. K., Kambanaros, M., and Katsos, N. (2016). The effect of childhood bilingualism and multilingualism on executive control. *Cognition* 149, 18–30. doi: 10.1016/j.cognition.2015.12.002
- Balladares, J., Marshall, C., and Griffiths, Y. (2016). Socio-economic status affects sentence repetition, but not non-word repetition, in Chilean preschoolers. *First Lang.* 36, 338–351. doi: 10.1177/0142723715626067
- Bernstein Ratner, N. (2000). "Elicited imitation and other methods for the analysis of trade-offs between speech and language skills in children," in *Methods for Studying Language Production*, eds L. Menn and N. Bernstein Ratner (Mahwah, NJ: Lawrence Erlbaum Associates), 291–311.
- Bishop, D. V. M. (2014). *Uncommon Understanding: Development and Disorders of Language Comprehension in Children*. Hove: Psychology Press.
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., and Greenhalgh, T. (2016). CATALISE: a multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLoS ONE* 11:e0158753. doi: 10.1371/journal.pone.0158753
- Bortolini, U., Arfé, B., Casello, C. M., Degasperi, L., Deevy, P., and Leonard, L. B. (2006). Clinical markers for specific language impairment in Italian: the contribution of clitics and non-word repetition. *Int. J. Lang. Commun. Disord.* 41, 695–712. doi: 10.1080/13682820600570831
- Bortolini, U., Caselli, M. C., Deevy, P., and Leonard, L. B. (2002). Specific language impairment in Italian: the first steps in the search for a clinical marker. *Int. J. Lang. Commun. Disord.* 37, 77–93. doi: 10.1080/13682820110116758
- Botting, N., and Conti-Ramsden, G. (2003). Autism, primary pragmatic difficulties, and specific language impairment: can we distinguish them using psycholinguistic markers? *Dev. Med. Child Neurol.* 45, 515–524.
- Boucher, J., and Lewis, V. (1997). *Preschool Language Scale—3*. London: Psychological Corporation.
- Briscoe, J., Bishop, D. V., and Norbury, C. F. (2001). Phonological processing, language, and literacy: a comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *J. Child Psychol. Psychiatry* 42, 329–340. doi: 10.1111/1469-7610.00726
- Brownwell, R. (2000). *Receptive One-Word Picture Vocabulary Test*. Novato, CA: Academic Therapy Publications.
- Chiat, S., and Roy, P. (2008). Early phonological and socio-cognitive skills as predictors of later language and social communication outcomes. *J. Child Psychol. Psychiatry* 49, 635–645. doi: 10.1111/j.1469-7610.2008.01881.x
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York, NY: Greenwood Publishing Group.
- Conti-Ramsden, G., Botting, N., and Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *J. Child Psychol. Psychiatry* 42, 741–748. doi: 10.1111/1469-7610.00770
- Devescovi, A., and Caselli, M. C. (2007). Sentence repetition as a measure of early grammatical development in Italian. *Int. J. Lang. Commun. Disord.* 42, 187–208. doi: 10.1080/13682820601030686
- Dockrell, J. E. (2001). Assessing language skills in preschool children. *Child Adolesc. Ment. Health* 6, 74–85. doi: 10.1111/1475-3588.00325
- Dockrell, J. E., and Marshall, C. R. (2015). Measurement issues: assessing language skills in young children. *Child Adolesc. Ment. Health* 20, 116–125. doi: 10.1111/camh.12072
- European Social Survey (2010). *Round 5 Source Showcards*. London: Centre for Comparative Social Surveys, City University London.
- Everitt, A. (2009). *Assessing the Potential of Risk Markers in Late Talkers to Identify Subsequent Specific Language Impairment*. Ph.D. dissertation, University of Aberdeen.
- Fey, M. E., and Cleave, P. L. (2008). Early language intervention. *Semin. Speech Lang.* 11, 165–181. doi: 10.1055/s-2008-1064250
- Friedmann, N., and Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: a study of SLI and normal development. *J. Child Lang.* 31, 661–681. doi: 10.1017/S0305000904006269
- Friedmann, N., and Novogrodsky, R. (2008). "Subtypes of SLI: SySLI, PhoSLI, LeSLI, and PraSLI," in *Language Acquisition and Development: Proceedings of GALA 2007*, eds A. Gavarró and M. J. Freitas (Newcastle-upon-Tyne: Cambridge Scholars), 205–217.
- Gallagher, A. L., and Chiat, S. (2009). Evaluation of speech and language therapy interventions for pre-school children with specific language impairment: a comparison of outcomes following specialist intensive, nursery-based and no intervention. *Int. J. Lang. Commun. Disord.* 44, 616–638. doi: 10.1080/13682820802276658
- Gardner, H., Froud, K., McClelland, A., and van der Lely, H. K. (2006). Development of the Grammar and Phonology Screening (GAPS) test to assess key markers of specific language and literacy difficulties in young children. *Int. J. Lang. Commun. Disord.* 41, 513–540. doi: 10.1080/13682820500442644
- Grohmann, K. K. (2014a). "CAT research on object clitic placement: Where are we now?" in *Developments in the Acquisition of Clitics*, eds K. K. Grohmann and T. Neokleous (Newcastle-upon-Tyne: Cambridge Scholars Publishing), 1–40.
- Grohmann, K. K. (2014b). Towards comparative bilingualism. *Linguist. Approach. Biling.* 4, 337–342. doi: 10.1075/lab.4.3.06gro
- Grohmann, K. K., and Kambanaros, M. (2016). The gradience of multilingualism in typical and impaired language development: positioning bilingualism within comparative bilingualism. *Front. Psychol.* 7:37. doi: 10.3389/fpsyg.2016.00037
- Grohmann, K. K., Papadopoulou, E., and Themistocleous, C. (2017). Acquiring clitic placement in bilingual settings: Interactions between social factors. *Front. Commun.* 2:5. doi: 10.3389/fcomm.2017.00005
- Kambanaros, M., and Grohmann, K. K. (2013). "Profiling (specific) language impairment in bilingual children: Preliminary evidence from Cyprus," in *Solutions in the Assessment of Bilinguals*, eds V. C. Mueller Gathercole (Bristol: Multilingual Matters), 146–174.
- Kambanaros, M., Grohmann, K. K., Michaelides, M., and Theodorou, E. (2013). Comparing multilingual children with SLI to their bilingual peers: evidence from object and action picture naming. *Int. J. Multiling.* 10, 60–81. doi: 10.1080/14790718.2012.705846
- Kamhi, A. G., Catts, H. W., and Davis, M. K. (1984). Management of sentence production demands. *J. Speech Hear. Res.* 27, 329–338. doi: 10.1044/jshr.2703.329
- Komeili, M., and Marshall, C. R. (2013). Sentence repetition as a measure of morphosyntax in monolingual and bilingual children. *Clin. Linguist. Phon.* 27, 152–162. doi: 10.3109/02699206.2012.751625

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02104/full#supplementary-material>

- Kunnari, S., Savinainen-Makkonen, T., Leonard, L. B., Mäkinen, L., and Tolonen, A.-K. (2014). The use of negative inflections by Finnish-speaking children with and without specific language impairment. *Clin. Linguist. Phon.* 28, 697–708. doi: 10.3109/02699206.2014.886725
- Leclercq, A. L., Quémart, P., Magis, D., and Maillart, C. (2014). The sentence repetition task: a powerful diagnostic tool for French children with specific language impairment. *Res. Dev. Disabil.* 35, 3423–3430. doi: 10.1016/j.ridd.2014.08.026
- Leivada, E., Kambanaros, M., Taxitari, L., and Grohmann, K. K. (2017). (Meta)linguistic abilities of bilingual educators: the case of Cyprus. *Int. J. Biling. Educ. Biling.* doi: 10.1080/13670050.2017.1401040. [Epub ahead of print].
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York, NY: Wiley.
- Leonard, L. B. (1998). *Children with Specific Language Impairment*. Cambridge, MA: MIT Press.
- Leonard, L. B. (2001). Fillers across languages and language abilities. *J. Child Lang.* 28, 257–261. doi: 10.1017/S0305000900004499
- Lust, B., Flynn, S., and Foley, C. (1996). “What children know about what they say: elicited imitation as a research method for assessing children’s syntax,” in *Methods for Assessing Children’s Syntax*, eds D. McDaniel, C. McKee, and S. Cairns Helen (Cambridge, MA: MIT Press), 55–76.
- Marinis, T., and Armon-Lotem, S. (2015). “Sentence repetition,” in *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. de Jong, and N. Meir (Bristol: Multilingual Matters), 95–122.
- Mastropavlou, M., and Tsimpli, I. M. (2011). Complementizers and subordination in typical language acquisition and SLI. *Lingua* 121, 442–462. doi: 10.1016/j.lingua.2010.10.009
- Newcomer, P. L., and Hammill, P. A. (1997). *Test of Language Development*. Austin, TX: PRO-ED.
- Newton, B. (1972). *Cypriot Greek: Its Phonology and Inflections*. The Hague: Mouton.
- Novogrodsky, R., and Friedmann, N. (2006). The production of relative clauses in syntactic SLI: a window to the nature of the impairment. *Int. J. Speech Lang. Pathol.* 8, 364–375. doi: 10.1080/14417040600919496
- Oetting, J. B., McDonald, J. L., Seidel, C. M., and Hegarty, M. (2016). Sentence recall by children with SLI across two mainstream dialects of English. *J. Speech Hear. Res.* 59, 183–194. doi: 10.1044/2015\_JSLHR-L-15-0036
- Plante, E., and Vance, R. (1994). Selection of preschool language tests: a data-based approach. *Lang. Speech Hear. Serv. Sch.* 25, 15–24. doi: 10.1044/0161-1461.2501.15
- Poljšenská, K., Chiat, S., and Roy, P. (2015). Sentence repetition: what does the task measure? *Int. J. Lang. Commun. Disord.* 50, 106–118. doi: 10.1111/1460-6984.12126
- Poll, G. H., Betz, S. K., and Miller, C. A. (2010). Identification of clinical markers of specific language impairment in adults. *J. Speech Hear. Res.* 53, 414–429. doi: 10.1044/1092-4388(2009/08-0016)
- Raven, J. C., Raven, J., and Court, J. (1998). *Raven’s Progressive Matrices*. San Antonio, TX: Harcourt Assessment.
- Redmond, S. M. (2005). Differentiating SLI from ADHD using children’s sentence recall and production of past tense morphology. *Clin. Linguist. Phon.* 19, 109–127. doi: 10.1080/02699200410001669870
- Redmond, S. M., Thompson, H. L., and Goldstein, S. (2011). Psycholinguistic profiling differentiates specific language impairment from typical development and from attention-deficit/hyperactivity disorder. *J. Speech Hear. Res.* 54, 99–117. doi: 10.1044/1092-4388(2010/10-0010)
- Riches, N. G., Loucas, T., Baird, G., Charman, T., and Simonoff, E. (2010). Sentence repetition in adolescents with specific language impairments and autism: an investigation of complex syntax. *Int. J. Lang. Commun. Disord.* 45, 47–60. doi: 10.3109/13682820802647676
- Rispens, J. E. (2004). *Syntactic and Phonological Processing in Developmental Dyslexia*. Ph.D. dissertation, Rijksuniversiteit Groningen.
- Rowe, C., and Grohmann, K. K. (2013). Discrete bilingualism: towards co-overt prestige and diglossic shift in Cyprus. *Int. J. Soc. Lang.* 224, 119–142. doi: 10.1515/ijsl-2013-0058
- Roy, P., Chiat, S., and Dodd, B. (2014). *Language and Socioeconomic Disadvantage: From Research to Practice*. London: City University London.
- Seeff-Gabriel, B., Chiat, S., and Dodd, B. (2010). Sentence imitation as a tool in identifying expressive morphosyntactic difficulties. *Int. J. Lang. Commun. Disord.* 45, 691–702. doi: 10.3109/13682820903509432
- Semel, E., Wiig, E., and Secord, W. (1989). *CELF-R Screening Test*. San Antonio, TX: Psychological Corporation.
- Sideridis, G., Antoniou, F., Mouzaki, A., and Simos, P. G. (2015). *Raven’s Coloured Progressive Matrices and Vocabulary [in Greek]*. Athina: Motivo.
- Sophocleous, A. (2011). Two languages in the classroom: the inconsistency between national and local objectives in formal education in Cyprus. *J. Lang. Identity Educ.* 10, 266–281. doi: 10.1080/15348458.2011.598129
- Stavrakaki, S. (2001). *Specific Language Impairment in Greek: Aspects of Syntactic Production and Comprehension*. Ph.D. dissertation, Aristotelian University of Thessaloniki.
- Stokes, S. F., Wong, A. M., Fletcher, P., and Leonard, L. B. (2006). Non word repetition and sentence repetition as clinical markers of specific language impairment: the case of Cantonese. *J. Speech Hear. Res.* 49, 219–236. doi: 10.1044/1092-4388(2006/019)
- Taxitari, L., Kambanaros, M., and Grohmann, K. K. (2015). A Cypriot Greek adaptation of the CDI: early production of translation equivalents in a bi(dia)lectal context. *J. Greek Linguist.* 15, 122–145. doi: 10.1163/15699846-01501003
- Taxitari, L., Kambanaros, M., Floros, G., and Grohmann, K. K. (2017). “Early language development in a bilingual context: the Cypriot adaptation of the MacArthur-Bates CDI,” in *Crosslinguistic Encounters with Language Acquisition: Typical and Atypical Development*, eds E. Babatsouli, D. Ingram, and N. Müller (Bristol: Multilingual Matters), 145–174.
- Theodorou, E. (2013). *Specific Language Impairment in CYPRIOT Greek: Diagnostic and Experimental Investigations*. Ph.D. dissertation, University of Cyprus, Nicosia.
- Theodorou, E., and Grohmann, K. K. (2012). The acquisition of relative clauses in Cypriot Greek: production and comprehension. *Diacritica* 26, 271–300.
- Theodorou, E., and Grohmann, K. K. (2015). Object clitics in Cypriot Greek children with SLI. *Lingua* 161, 144–158. doi: 10.1016/j.lingua.2014.11.011
- Theodorou, E., Kambanaros, M., and Grohmann, K. K. (2013). Specific language impairment: diagnostic issues. *Lang. Variat.* 13, 217–236. doi: 10.1075/lv.13.2.04the
- Theodorou, E., Kambanaros, M., and Grohmann, K. K. (2016). Diagnosing bilingual children with SLI: determination of identification accuracy. *Clin. Linguist. Phon.* 30, 925–931. doi: 10.1080/02699206.2016.1182591
- Thordardottir, E. (2015). The relationship between bilingual exposure and morphosyntactic development. *Int. J. Speech Lang. Pathol.* 17, 97–114. doi: 10.3109/17549507.2014.923509
- Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., et al. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *J. Speech. Lang. Hear. Res.* 54, 580–597. doi: 10.1044/1092-4388(2010/09-0196)
- van der Lely, H. K. (2003). “Do heterogeneous SLI deficits need heterogeneous theories? SLI subgroups, G-SLI and the RDDR hypothesis,” in *Towards a Definition of Specific Language Impairment*, eds Y. Levy and J. Schaeffer (Lawrence Erlbaum), 109–134.
- van der Lely, H. K. J., and Battell, J. (2003). Wh-movement in children with grammatical SLI: a test of the RDDR Hypothesis. *Language* 79, 153–181. doi: 10.1353/lan.2003.0089
- Vender, C., Borgia, R., Cumer Bruno, S., Freo, P., and Cardini, G. (1981). Un test di ripetizione di frasi: analisi delle performances in bambini normali. *Giornale de Neuropsichiatria Infantile* 24, 819–831.
- Vinther, T. (2002). Elicited imitation: a brief overview. *Int. J. App. Linguist.* 12, 54–73. doi: 10.1111/1473-4192.00024
- Wiig, E. H., Secord, W., and Semel, E. (1992). *Clinical Evaluation of Language Fundamentals – Preschool*. San Antonio, TX: Psychological Corporation.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Theodorou, Kambanaros and Grohmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Length of Utterance, in Morphemes or in Words?: MLU3-w, a Reliable Measure of Language Development in Early Basque

Maria-José Ezeizabarrena<sup>1\*</sup> and Iñaki Garcia Fernandez<sup>2</sup>

<sup>1</sup> Department of Linguistics and Basque Studies, Faculty of Arts, University of the Basque Country, Vitoria/Gasteiz, Spain,

<sup>2</sup> Department of Social Psychology and Methodology of Behavioral Sciences, Faculty of Psychology, University of the Basque Country, Donostia-San Sebastian, Spain

## OPEN ACCESS

### Edited by:

Aritz Irurtzun,  
Centre national de la recherche  
scientifique (CNRS), France

### Reviewed by:

Lluís Barceló-Coblijn,  
University of the Balearic Islands,  
Spain  
Marilyn Vihman,  
University of York, United Kingdom

### \*Correspondence:

Maria-José Ezeizabarrena  
mj.ezeizabarrena@ehu.es

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 29 July 2017

**Accepted:** 13 December 2017

**Published:** 08 January 2018

### Citation:

Ezeizabarrena M-J and Garcia  
Fernandez I (2018) Length of  
Utterance, in Morphemes or in  
Words?: MLU3-w, a Reliable Measure  
of Language Development in Early  
Basque. *Front. Psychol.* 8:2265.  
doi: 10.3389/fpsyg.2017.02265

The mean length of utterance (MLU), which was proposed by Brown (1973) as a better index for language development in children than age, has been regularly reported in case studies as well as in cross-sectional studies on early spontaneous language production. Despite the reliability of MLU as a measure of (morpho-)syntactic development having been called into question, its extensive use in language acquisition studies highlights its utility not only for intra- and inter-individual comparison in monolingual language acquisition, but also for cross-linguistic assessment and comparison of bilinguals' early language development (Müller, 1993; Yip and Matthews, 2006; Meisel, 2011). An additional issue concerns whether MLU should be measured in words (MLU-w) or morphemes (MLU-m), the latter option being the most difficult to gauge, since new challenges have arisen regarding how to count zero morphemes, suppletive and fused morphemes. The different criteria have consequences, especially when comparing development in languages with diverging morphological complexity. A variant of MLU, the MLU3, which is calculated out of the three longest sentences produced (MLU3-w and MLU3-m), is included among the subscales of expressive language development in CDI parental reports (Fenson et al., 1993, 2007). The aim of the study is to investigate the consistency and utility of MLU3-w and MLU3-m as a measure for (morpho-)syntactic development in Basque, an agglutinative language. To that end, cross-sectional data were obtained using either the Basque CDI-2 instrument (16- to 30-month-olds) or the Basque CDI-3 (30- to 50-month-olds). The results of analyzing reports on over 1,200 children show three main findings. First, MLU3-w and MLU3-m can report equally well on very young children's development. Second, the strong correlations found between MLU3 and expressive vocabulary in the Basque CDI-2 and CDI-3 instruments, as well as between MLU3 and both nominal and verbal morphology scales, confirm the consistency not only of MLU3 but also of the two Basque CDI instruments. Finally, both MLU3-w and MLU3-m subscales appear sensitive to input after age 2, which emphasizes their utility for identifying developmental patterns in Basque bilinguals.

**Keywords:** MLU, Basque language, early language development, bilingualism, complexity

## INTRODUCTION

### Mean Length of Utterance: MLU and MLU3

How to measure language complexity is a question that has occupied linguists in a longstanding debate. Some authors maintain that since all languages are learnable by any child, they must have the same degree of complexity. In this regard, cross-linguistic differences found in complexity in each language component are believed to be the result of a compensation system, so that languages showing very high complexity in one particular domain are expected to have less complexity in other domains and vice-versa. In addition, the observation that, synchronically, many languages with low complexity in morphology have a rigid word order or a more complex phonological system than languages with complex morphology may support that assumption. However, counter-evidence has also been provided by scholars denying any theory-internal reason to predict similar degrees of complexity in all natural languages. See Newmeyer (2017) and Newmeyer and Preston (2014) for an overview of the debate.

The issue of language complexity piqued early language acquisition researchers' interest already in the beginning of the twentieth century. Such is the case of, for example, Nice (1925), who regarded *average sentence length* as "the most important single criterion for judging a child's progress in the attainment of adult language" (Rice et al., 2010). In a similar vein, five decades later, Roger Brown passionately defended his *Mean Length of Utterance* or *MLU*, which proved to be one of the most commonly-mentioned indexes of constructional complexity in child language by the end of the century:

"... The MLU is an excellent simple index of grammatical development because almost every new kind of knowledge increases length: the number of semantic roles expressed in a sentence, the addition of obligatory morphemes, coding modulation of meaning [...]. Jand, of course, embedding and coordinating. All alike have the common effect on the surface form of the sentence of increasing length (especially if measured in morphemes, which includes bound forms like inflections rather than words)" (Brown, 1973, pp. 53–54).

Brown considered MLU to be a more suitable index than age to compare individuals' development, since it permits identifying "on internal grounds" children who are "at the same level of constructional complexity" but who may not be "of the same chronological age" (Brown, 1973, p. 55).

In addition to the MLU calculated from the sentence sample uttered in a recording session, Brown regarded the upper bound or the longest sentence produced at a specific age as a relevant additional index to measure the attained grammar complexity of children. Thus, he established a sequence of five stages in children's earliest morphosyntactic development based on the two indexes: MLU and *upper bound*. Both values increased with age in the three longitudinal corpora analyzed (Eve, Adam, and Sarah). Each stage was associated with the child's productive use (at least in 90% of the contexts in which they are required) of some linguistic structures, and individual differences were observed in the age at which each child reached the various stages. For instance, Eve attained stage V at 2;2

years, whilst at that age Adam's and Sarah's MLU values around 2 indicated stage II. In **Table 1** we have combined data which Brown presented separately: the target values of MLU and upper bound corresponding to each stage and the age ranges of the three children studied longitudinally at the different stages. The variability in age is evidenced by the large age ranges across stages displayed in column 4.

Despite the advantages of an index other than age to compare children's linguistic development, Brown still pointed out some limitations, starting from Stage V onwards. He argued that, at that stage, children's varied linguistic productions and their MLU begin to depend more on the nature of the interaction than on what children know (Brown, 1973, p. 54).

Brown's view of complexity is not related to any specific language component such as semantics or morphology. It is based on the assumption that the acquisition of components such as  $x$  and  $y$  alone does not immediately, or even relatively quickly, lead to the acquisition of the construction  $x + y$  that combines the two. Consequently, in his *cummulative* sense of complexity, "construction  $x + y$  may be regarded as more complex than  $x$  or  $y$  because it involves everything involved in either of the constructions alone, plus something more" (Brown, 1973, p. 400). This lack of precision is probably what led researchers to question MLU's appropriateness to measure *morphosyntactic development*. Bickerton (1991), for instance, suggested that qualitative aspects of syntactic development cannot be directly evaluated, since the increase in length of utterances does not necessarily imply an increase in syntactic complexity. In fact, similar or higher MLU values (1a-c) may correspond to utterances with a lower morphosyntactic complexity, which is the case with the coordinated structures in (1a) as compared to S-V agreement examples in (1b) or the embedding structures in (1c).

- (1) a. *Peter and Mary* (3 w / 3 m)
- b. *Ann comes* (2 w / 3 m)
- c. *want to come* (3 w / 3 m)

Thus, MLU may appear to be a quantitative rather than a qualitative measurement: "as utterances get longer and MLU

**TABLE 1** | Target values and approximations attained for MLU and upper bounds.

Stage	Target value		Age in months of Adam, Eve, and Sarah
	MLU	Upper bound	
I Semantic Roles and Syntactic Relations	1.75	5	18–30
II Grammatical Morphemes and the Modulation of Meaning	2.25	7	19–34
III Modalities of the Simple Sentence	2.75	9	20–36
IV Embedding	3.50	11	22–42
V	4.00	13	24–48

Brown (1973, pp. 56–57).



increases, some sort of increase in complexity is bound to occur, but there is no *a priori* reason why the increase should take only the forms it does, and, in particular, that these forms should be the same for all children studied, whatever the language in question” (Brown, 1973, pp. 64–65). Additionally, issues such as how to measure children’s achieved linguistic complexity and whether the same degree of complexity should be assumed at a particular stage cross-linguistically or across individuals acquiring a particular language have not received a convincing and generally accepted answer yet.

However, the generalized acquisition order of 14 inflectional markers in English established by Brown, which was confirmed in later longitudinal studies, reinforces the supposition of some pattern in morphosyntactic development which goes beyond the aforementioned individual variability. Despite MLU being originally “invented for English,” Brown was still aware of its utility in other languages for cross-linguistic comparison, once some adjustments were made: “Studies of highly inflected languages [...], all report some difficulty in adapting our rules of calculation, invented for English, which is minimally inflected, to their languages. What I have used is, in each case, the author’s choice of the linguistically most reasonable value” (Brown, 1973, p. 68). Actually, many longitudinal case studies conducted in typologically distant languages have provided relevant results regarding the specific structures which arise in children’s spontaneous production at each specific developmental stage. Besides, MLU has been used in cross-sectional studies comparing early bilingual children’s development in their two languages (Marchman et al., 2004; Meisel, 2011; Thordardottir, 2011; Hoff et al., 2014) as well as typical vs. atypical language development (Johnston, 2001; Rice et al., 2010; Wiczorek, 2010).

In his seminal 1973 book, Brown devoted part of the introductory section to describing and discussing the set of rules for calculating MLU and upper bound in spontaneous production corpora. Here are the most relevant ones: (a) a subsample is required to calculate MLU in a longer sample gathered at some specific developmental stage. However, not every utterance can be equally reliable in the sample: 100 utterances should be taken from the fully transcribed utterances, starting at the second transcription page rather than from the first minutes of the conversation; (b) stuttering or repeated attempts to produce some words or utterances are counted once, in the most complete form used. This rule may avoid under-scoring due to the selection of non-representative items of the child’s (real) linguistic performance in constructional complexity; (c) fillers such as *umm* are not counted, in contrast to *no*, *yeah*, *hi*, which are included in the counting; (d) inflectional morphemes (plural, genitive, 3rd singular present *-s*, and so on) are counted as separate morphemes and inflected auxiliaries are counted as mono-morphemic words, as are compounds, for example, *birthday*. In our opinion, such counting criteria appear as an intermediate option between counting words and morphemes. However, such a counting system, together with the specific properties of English morphosyntax (a limited inventory of inflectional person and plural markers, low word complexity) and the scarcity of inflectional markers in children’s early productions, may lead one to predict no great difference

in measuring English child utterance length in words or in morphemes. In contrast, in languages with a certain degree of morphological complexity, like Basque, many researchers are in favor of measuring morphosyntactic development in morphemes rather than in words (Idiazabal, 1991; Barreña, 1995; Ezeizabarrena, 1996; Elosegi, 1998; Larrañaga, 2000; Larrañaga and Guijarro-Fuentes, 2012a). Nonetheless, the high (almost perfect) intralinguistic correlations between the two ways of calculating MLU found in such typologically distant languages as Spanish (Aguado, 1995; Jackson-Maldonado and Conboy, 2007), Irish, Icelandic and Dutch (see Parker and Brorson, 2005 and references therein), indicates that MLU-m may not necessarily be a better measurement than MLU-w. In contrast to authors who have suggested the higher usefulness of MLU-w because of the ease of calculating it, Wiczorek (2010) has questioned the fact that MLU-w and MLU-m can be regarded as similar indicators of morphosyntactic development simply because of the high correlations attested cross-linguistically. According to this researcher, MLU-w is related to lexical development rather than to grammatical development and therefore, the opposite is expected to be the case for MLU-m, which should show a stronger relation to grammatical rather than lexical development. A third way of calculating MLU in syllables (MLU-s) has also been explored in Irish (Hickey, 1991) and in Inuktitut (Allen and Dench, 2015). Surprisingly, MLU-s, which *a priori* would not be considered an index of grammatical development *per se*, or at least not in every language, also correlates with the previous indexes. The high correlations attested across languages between the different types of MLU may indirectly cast doubt on the “equivalence” of all of them as measures of language development, although determining exactly what the different variants of MLU measure in each language goes far beyond the aim of the current study.

Apart from the several ways of counting MLU, another objection to the use of MLU is the subjectivity present throughout the different steps preceding its calculation. To start with, MLU is sensitive to event and exchange patterns, situational variability and conversational dominance in a bilingual child, which may cause the sample collection on a particular date or conversational situation not to be the best example of the child’s regular linguistic use (see Johnston, 2001 and references therein). Thus, counting all the sentences in a session or selecting the (50?, 100?, more?) utterances from the first, intermediate or final part of a two-hour recorded conversation may result in a different MLU value of a child’s production at a particular age. Moreover, criteria for calculating MLU vary across studies, such as in the case of *MLU* vs. *alternate MLU* measures (Johnston, 2001), or of measuring MLU in words (MLU-w), morphemes (MLU-m) or syllables (MLU-s). Finally, subjectivity is present in the process of transcribing and coding oral speech in general, a task which “relies on the accuracy of the transcriber” (Rollins et al., 1996) and in the process of segmenting utterances. Segmenting words and especially morphemes in an utterance arises as the next complication in the process, where decisions regarding null morphemes, multimorphemic words such as portmanteaux, compounds and so on need to be made before starting with the analysis. Otherwise the variability

found in children's spontaneous productions may lead to quite diverging value assignments to the same utterance. In order to regulate the subjectivity inherent in the processes mentioned above, single *individuals are put in charge* of the segmentation task of a whole set of recordings or of a sample collection, and further *interjudge reliability* rates are established on their codifications.

Despite the objections discussed earlier, MLU has still been extensively used in both intra- and inter-individual comparative studies. This is the case of, for instance, studies on language dominance which compare bilinguals' development in their two languages. On the assumption that length of utterances across languages may vary more depending on the unit in which its calculation is based, MLU-m has been proposed as a better measure for bilinguals' individual interlinguistic comparison in language pairs such as Basque-Spanish (Meisel, 1994; Ezeizabarrena, 1996; Elosegi, 1998; Larrañaga, 2000; Larrañaga and Guijarro-Fuentes, 2012a etc.), whilst studies on French-German bilinguals (Meisel, 1991; Müller, 1993; Müller and Kupisch, 2003; Kupisch, 2008; Schmeiser et al., 2016) or English-Mandarin bilinguals (Yip and Matthews, 2006) and even some on Spanish-Basque (Larrañaga and Guijarro-Fuentes, 2012b) have opted for MLU-w. See also Hickey (1991), who considers that MLU's utility for cross-linguistic comparison cannot be generalized even intraindividually.

Despite criticisms, MLU, in its different modalities, remains as one very relevant index for morphosyntactic development in longitudinal corpora of spontaneous language production, and the inclusion of some versions of it in assessment instruments confirms this fact. Such is the case of *MLU3*, included in the MacArthur-Bates Communicative Development Inventories (CDI) instrument (Fenson et al., 1993, 2007), a parental questionnaire designed to obtain normative data which may allow researchers to assess both typically and atypically developing children. The *MLU3* is a combination of two indexes on which Brown's 5-stage classification was based (*mean length of utterance* and the *upper bound*). Yet *MLU3* has the particularity that the mean length is calculated based on the child's three longest recently-produced sentences according to their parents, instead of on a specific sample of child utterances gauged by a researcher in a longitudinal corpus.

Studies on early bilingualism using this measurement have concluded that *MLU3* values are sensitive to the amount of a child's exposure to the language. Bilinguals, who by definition have less exposure to their language(s) than monolinguals, have shown lower values than their age-matched monolingual counterparts (1;10–2;6: Hoff et al., 2012, 2014). More specifically, the results from Spanish-English bilingual groups, which were distinguished according to their higher, balanced and lower exposure to the language, revealed that the less input bilinguals had received in the language under study, the lower the scores they obtained in *MLU3* values (Hoff et al., 2012).

## Utterance Length in Basque

From the genetic point of view, Basque is unrelated to any other known language; that is, it is an isolate language. Typologically,

Basque is a null subject, ergative language with non-rigid SOV word order, a language with very rich nominal and verbal inflection (case marking, person and number subject-, direct object- and indirect object-agreement marking in the verb), with a predominantly agglutinative morphology and affixed postpositions. As a result, most nominal and verbal words comprise two or more morphemes (2a-c), which makes utterance length diverge, depending on whether it is measured in words (1,1 and 4 w) or morphemes (2, 4, and 8 m) in (2a), (2b) and (2c), respectively.

- (2) a. *panpin-a*  
doll-Det (1 w, 2 m)  
'doll' or 'the doll'
- b. *panpin-txo-a-rekin*  
doll-DIM-Det-with (1 w, 4 m)  
'with the dolly'
- c. *Jon panpin-txo-a-rekin etorri-ko da*  
Jon doll-DIM-the-with come-FU Aux.S3s (4 w, 8 m)  
'Jon will come with the dolly'

However, not all morphemes are counted as productive morphology in early child productions. Following Brown's (1973) proposal of counting productive (non-rote learned) words and morphemes and taking into account both the specific morphosyntactic properties, as well as the characteristics of earliest productions in Basque, Idiazabal (1991) established the first list of rules to calculate *MLU-m* in Basque, which were followed in later longitudinal case studies (Barreña, 1995; Ezeizabarrena, 1996; Elosegi, 1998; Almgren, 2000; Larrañaga, 2000). According to these rules, diminutive suffix *-txo* is not counted as a morpheme in very frequent diminutive words in child and child-directed speech such as *ama-txo* "mumm-y" and *aita-txo* "dadd-y" (1 w / 1 m) but, on the other hand, *-txo* is counted as a morpheme in the rest of the few remaining words that include it (2a-c). Moreover, the  $-\emptyset$  morpheme is not counted, and the *-a* ending, which is translated as Det(erminer) in the (2a, 2b) glosses, is not counted as a morpheme either. There are several reasons for not counting this *-a* ending, which is suffixed to the nominal phrase rather than to the noun, as a (productive) morpheme: (a) many lexical roots having an organic *-a* ending do not modify their phonology when the determiner *-a* is suffixed (*musika* "music/music-Det"), (b) overtly determined roots like *etxe-a* "house-Det" cannot always be considered as such, since they can be used to respond to the question, "how do you say... *house* in Basque?", where no determined nouns are expected; and (c) in early child Basque the nominal *-a* ending acts as an unanalyzed word boundary, rather than as a grammatical element, as seen in examples like *beste a umea* instead of *beste umea* "other child," attested in several longitudinal samples (Barreña and Ezeizabarrena, 1999).

## Sociolinguistic Context

Basque is a language spoken in the North Eastern area of Spain and the South West area of France, on both sides of the Atlantic Pyrenean mountains. All adult speakers of that language are bilingual Spanish-Basque or French-Basque. The

Basque-speaking community of roughly one million speakers mostly comprises people who grew up in Basque-speaking families and acquired Basque as their L1 (either simultaneously or alongside Spanish or French, successively) and early L2 speakers who, growing up in almost monolingual Spanish or French families, are exposed to Basque very early (from age 2 or 3 onwards) through the educational system. Another group of late L2 speakers acquired that language through adult training courses. Sociolinguistic surveys conducted in 2006 with population older than 15 years of age in the Basque Country described the following distribution of linguistic profiles: 15.4% passive bilinguals, 25.7% active bilinguals and 58.9% French or Spanish monolinguals. Further censal surveys conducted in the Basque Autonomous Community, the region in which most of the current sample was collected, concluded that 39% of the 5- to 9-year-old population had Basque exclusively or together with Spanish as their home language (Basque Government, 2009). Consequently, most L1 Basque-speaking children are exposed to different degrees of Spanish (or French) input, and this is also the case of the participants of our study.

## Aims and Predictions

The current paper investigates MLU3 scales' reliability as compared to other scales of the Basque CDI to assess early language development in that agglutinative language. For that, it provides data of 16- to 50-month-old children obtained using the Basque versions of the MacArthur-Bates CDI parental questionnaires.

In a language community such as the Basque-speaking one, in which being bilingual is the norm rather than the exception, the assumption that monolingual data are the best reference for "typical development" does not hold, and consequently, only instruments which are sensitive to the amount of exposure to the language(s) can accurately assess early bilingual language development. Therefore, a further study conducted with a subsample of over 1200 18- to 48-month-olds' MLU3-w and MLU3-m scores will analyse those measurements' sensitivity to two variables, chronological age and (relative) amount of exposure to the Basque language, with the aim of checking MLU3 subscales' utility in that particular context. Three predictions can be stated in this regard:

1. MLU3 scales will be as sensitive as the rest of the scales in the Basque CDI instrument to detect children's developmental changes as found in previous studies, and will reflect development in morphological complexity (Fenson et al., 1993, 2007).
2. Taking into account the morphosyntactic properties of an agglutinative language with rich morphology, such as Basque, MLU3 measured in morphemes will prove to be more discriminative than the MLU3 measured in words.
3. Input quantity will affect children's expressive language. Hence, differences in length of utterance are expected among bilinguals, depending on children's relative amount of exposure to Basque, as widely reported in early bilingual research (Marchman et al., 2004; Meisel, 2011; Thordardottir, 2011; Hoff et al., 2014).

## MATERIALS AND METHODS

### Instruments

The MacArthur-Bates Communicative Development Inventory (CDI) instrument is a parental questionnaire used to gather information regarding children's language use. Different versions of the instrument have been developed, all designed for different age ranges (CDI-1 for 8–15 months, CDI-2 for 16–30 months, and CDI-3 for 30–50 months) and for different purposes such as screening (short CDI-1 and CDI-2) or clinical diagnosis and research (full CDI-1 and CDI-2 questionnaires) (see Fenson et al., 2000). The CDI-1 is the only instrument which includes vocabulary comprehension in addition to expressive vocabulary and grammar. In contrast, CDI-2 and CDI-3 are oriented to expressive language use.

The current study reports on data obtained with the long version of the CDI-2 and the CDI-3, for which there is only one (short) version. The Basque version of the full CDI-2 instrument (16–30 months), henceforth *BCDI-2*, contains different sections such as vocabulary and morphology, in which informants tick the items their child already produces, some questions about whether the child has started combining words, as well as a section for writing down the child's three longest recently-produced sentences. In addition, there is a list of multiple-choice items in which informants choose, from the different options the one that best fits with the child's current production. Filling in this questionnaire may take between 10' and 60', depending on the child's level of expressive use.

The Basque version of the CDI-3 instrument (30–50 months), henceforth the *BCDI-3*, is much shorter than the CDI-2. The *BCDI-3* contains a vocabulary list, a grammar section, a section for writing down the three longest utterances, a list of multiple-choice items and a list of questions intended to assess children's knowledge of some logical and mathematical terms.

The sections and number of items analyzed in the current study are presented in **Table 2**. Neither the 37/29 items of the multiple-choice item section nor the 10 yes/no questions on logical concepts (included only in *BCDI-3*) have been included in the current analysis, since they are less homogeneous in format, across items and across the two instruments.

### Participants

The parents of over 2,000 children aged between 16 and 50 months of age participated in the study, filling in one of the

**TABLE 2 |** Number of items in the BCDI scales included in the study.

	BCDI-2	BCDI-3
Vocabulary	643 <sup>a</sup>	120
Nominal morphology	17	16
Verbal morphology	40	20
MLU3		

<sup>a</sup>For the current study, some postpositions, included in the vocabulary section of the questionnaire were analyzed as morphological suffixes rather than as vocabulary items. Consequently, the distribution of (vocabulary/grammatical) items included in this study will vary from previous studies such as Barreña et al.'s (2008a,b), conducted with the same data sample.

two instruments: either the BCDI-2 (16–30 months) instrument (Barreña et al., 2008a) or the BCDI-3 (30–50 months) instrument (Garcia et al., 2014). The questionnaire is written exclusively in Basque. Consequently, all the informants in this study are bilingual parents with different levels of language use who interact in Basque and (at least) one other language on a daily basis and address their child (some exclusively, others mostly or only sometimes) in Basque. Participants gave informed consent prior to participation. The study was approved by the ethics commission of the University of the Basque Country.

The data sampling lasted over a decade. The initial data collection of 2,248 questionnaires (BCDI-2  $n = 1,204$  / BCDI-3  $n = 1,044$ ) was filtered out based on a set of exclusion criteria: out of the age range (101 out of 15–30 months/26 older than 50 months), below 8-month-pregnancy pre-term born children (15/7), children who had over two ear infections during the first year (20/55); questionnaires in which vocabulary and/or grammar sections were incomplete (93/0) and questionnaires where any (one, two, or three) of the three longest utterances produced (207/389) and/or input data (25/15) were missing. Thus, the data sample of 16- to 50-month-olds analyzed for the current study includes 1,337 questionnaires (BCDI-2  $n = 750$ /BCDI-3  $n = 587$ ). As shown in **Figure 1**, all age groups (in months) consist of a range of 20–64 participants for the whole period studied. As for gender, girls and boys are evenly distributed across the age groups [ $\chi^2_{(14)} = 6.27$ ,  $p = 0.96$  in BCDI-2 and  $\chi^2_{(20)} = 28.18$ ,  $p = 0.11$  in BCDI-3].

In order to investigate the effect of input and age and the interaction between these two variables on MLU scores, the sample was limited to children aged between 1;6 and 4 years. The sub-sample of 1202 participants was divided into age groups and

input groups. Five groups resulted from the division in six-month age groups (18–24 months, 25–30 months, 31–36 months, 37–42 months and 43–48 months). Each age group was further divided into four different input groups based on the relative amount of exposure to Basque and Spanish: Monolingual or M (over 90% Basque input), Basque-dominant bilingual or BDB (Basque input 60–90%), Balanced Bilingual or BB (Basque input 40–60%) and Spanish-dominant bilingual or SDB (below 40% Basque input) (see **Table 3**). In what follows, we will use the terms *input* or *relative input* to refer to the relative amount of exposure to Basque and Spanish, following Thordardottir (2011), among others.

## Procedure and Coding

As in the original CDI, the grammar section of the BCDI includes several items regarding nominal inflection, verbal inflection and an item in which participants are requested to report on the child's longest three sentences produced recently. The MLU3 was calculated from the three utterances reported, as displayed in (3).

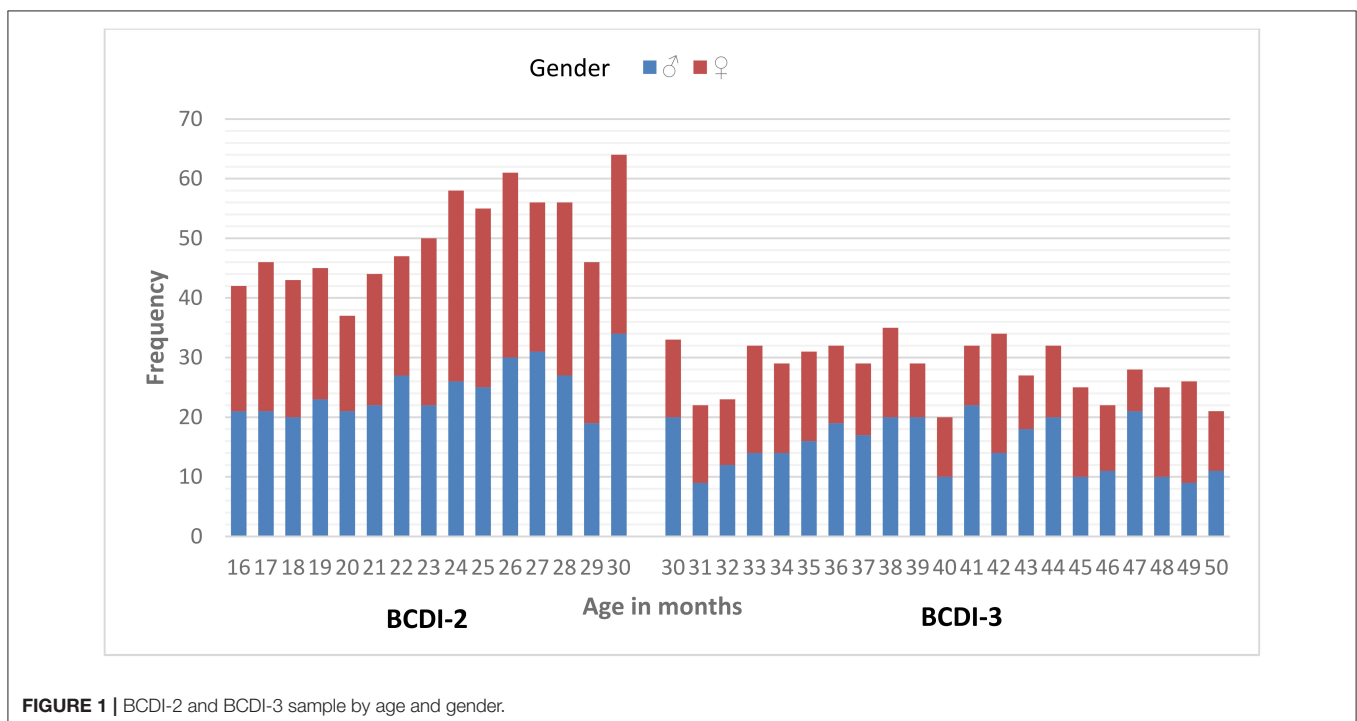
(3) Idatzi zure haurrak azken aldian esan dituen hiru esaldi luzeenak. 'Please write down the longest three sentences your child has recently produced':

a. Ni-k ur-a-Ø nahi du-t (4w 6m)  
I-Erg water-Det-Abs want Aux.S1s.O3s

a'. 'I want water' (3w 3m)

b. Zu-Ø kale-ra joan-Ø z-ea (4w 6m)  
you-Abs street-to go Aux.S2s

b'. 'you have gone/went to the street' (5w, 6m)



**FIGURE 1** | BCDI-2 and BCDI-3 sample by age and gender.

**TABLE 3** | Distribution of the sample (raw numbers of participants and percentages) in age and input groups.

Input groups	Age groups (range in months)				
	18–24 months	25–30 months	31–36 months	37–42 months	43–48 months
Monolingual	200 (61.7%)	221 (59.6%)	107 63.3(%)	109 (60.9%)	102 (64.1%)
Basque-dominant bilingual	63 (19.4%)	96 (25.9%)	26 (15.4%)	26 (14.5%)	26 (16.4%)
Balanced bilingual	41 (12.6%)	35 (9.4%)	22 (13.0%)	20 (11.2%)	19 (11.9%)
Spanish-dominant bilingual	20 (6.2%)	19 (5.1%)	14 (8.3%)	24 (13.4%)	12 (7.6%)
Total	324 (100%)	371 (100%)	169 (100%)	179 (100%)	159 (100%)

c. Unai-ren bila g-oa-z (3w 5m)  
Unai-Gen look-for go.S1pl

c'. 'we go looking-for Unai' (4w, 4m)

Examples in (3) illustrate the three longest utterances of a 28-month-old child randomly chosen from the BCDI-2 sample and the way they were measured. Thus, MLU3 in (3) was calculated based on the mean of the length of the three utterances reported. So that MLU3-w of (3a + 3b + 3c) / 3 is (4 + 4 + 3) / 3, that is, 3.66 and MLU3-m is (6 + 6 + 5) / 3, namely, 5.66. This shows that MLU-w and MLU3-m differ considerably in Basque. In contrast, measuring utterance length in MLU-w or MLU-m in a language with predominantly monomorphemic words like English (3a', 3b', 3c') does not make much difference: MLU3-w: 12/3 = 4; MLU3-m = 13/3 = 4.33.

MLU3-w and MLU3-m calculations were performed by two independent coders. The high coefficients of intraclass correlation resulting from the statistical analysis for both MLU3 scales in the two instruments ( $r = 0.91$  and  $\alpha = 0.95$  for MLU3-w;  $r = 0.94$  and  $\alpha = 0.96$  for MLU3-m in BCDI-2;  $r = 0.95$  and  $\alpha = 0.97$  for MLU3-w;  $r = 0.95$  and  $\alpha = 0.98$  for MLU3-m in BCDI-3) confirmed an excellent interjudge reliability of the data (Koo and Li, 2016).

Only the children who had not started combining words yet (their parents responded with “not yet” to the item preceding the three longest utterance section) obtained 1 as a mean value for the two variables, MLU3-w and MLU3-m. The rest of the children obtained higher values.

The results from the MLU sections will be analyzed together with the scores obtained in three more scales: vocabulary, nominal inflection, and verb morphology. In the vocabulary and morphology sections, informants were asked to tick the items their child had started producing. The final score was calculated by summing up the total number of items ticked in each of the sections.

The maximal potential score in *vocabulary* was 643 items in BCDI-2 and 120 in BCDI-3. *MLU3-w* and *MLU3-m* were open scales and therefore no maximal values could be estimated *a priori*.

As for *nominal morphological markers*, 17 items from BCDI-2 and 14 items from BCDI-3 were analyzed for the current study and consequently, 17/14 were the highest possible scores in this section, respectively. The items analyzed from BCDI-2 are the following: 11 postpositional suffixes (-n, -ra, -raino, -rantz, -tik, -zkoa, -koa, -z, -rena, -rentzat, -rekin) and 6 non-postpositional

ones (plural -k, genitive possessive -ren, genitive locative -ko, ergat -k, dative -ri, and diminutive -txo)<sup>1</sup>. BCDI-3 contains 11 postpositions (-n, -ra, -raino, -tik, -zkoa, -koa, -z, -rena, -rentzat, -rekin, -rengna) and 3 more nominal suffixes (plural -k, ergative -k, dative -ri).

As for *verbal inflection*, the maximal possible score was 39 in BCDI-2 and 22 in BCDI-3, corresponding to the number of items included in the two instruments in the current study. The items in BCDI-2 are three aspectual suffixes (imperfective -tzen, future -ko, and perfective -ta) in addition to 36 inflected frequent verb forms, most of them auxiliary forms. The items included in BCDI-3 (22) are two aspectual suffixes (imperfective -tzen, and future -ko) and 20 very frequent, most of them inflected auxiliary verb forms (*naiz* “am,” *da* “is,” *dago* “is,” *dizut* “I have... it to you,” *zenuen* “you had... it”).

## Data Analysis

One-way ANOVAs were conducted separately for BCDI-1 and BCDI-2 instruments in order to measure the effect of age. In addition, Pearson's correlations were calculated to analyse between-scale relations, and finally, partial correlation coefficients were computed between BCDI scales with age as the covariate.

On the other hand, two-way ANOVAs were performed to compare the main effects of age and input in the whole sample, as well as the interaction between age and input in MLU3-w and MLU3-m scales. The effect size was calculated according to Cohen (1992) and Richardson (2011).

## RESULTS

A variety of structures and morphological markers are attested in the sample of utterances produced by the participants, based on their parents' reports. The examples of 24-month-olds listed in (4a-b) and of 30-month-olds in (4c-d) were collected using the BCDI-2, whereas examples from 30-month-olds (4e-f), 36-month-olds (4f-h), 42-month olds (4i-j) and 48-month-olds (4k-l) were obtained using the BCDI-3 instrument. As expected in a language with rich case and inflectional morphology, length of utterance varies depending on whether it is measured in w(ords) or in m(orphemes) and the older the children become, the more complex are the structures attested. Thus, morphologically

<sup>1</sup>Only the very few instances of -txo attached to words other than *ama*, *amatxo* “mom, mommy” and *aita*, *aitatxo* “dad, daddy” were counted as morphemes.

complex structures which are rare among children younger than 30 months, such as inflected verb forms with multiple agreement markers (4d), postpositional complex phrases (4f, 4h), embedded sentences carrying embedding particles (9g, 9k, 9l), start being reported from 2;6 and 3 years onwards or even later.

- a. amona etxea-n dago (cod. 628, 24 months: 3 w / 4 m)  
Grandma house-in is  
'Grandma is at home'
- b. Josu-k apurtu dau (cod. 455, 24 months: 3 w / 5 m)  
Josu-Erg. break have.S3s  
'Josu has broken (something)'
- c. nahi duzu ni-rekin jolastu kale-an? (cod. 481, 30 months: 5 w / 8 m)  
want have.S2s I-with play street-in  
'Do you want to play with me on the street?'
- d. Ez uz-ten hau egi-ten (cod. 1110, 30 months: 4 w / 6 m)  
Neg permit-ting this do-ing  
'not permitting doing this'  
Intended: *ez dit uzten hau egiten* 'it does not permit me to do this'
- e. Ni bakarrik esnatu naiz (cod. 1041, 30 months, 4 w / 5 m)  
I alone wake be.1s  
'I woke up alone'
- f. Amatxu lan-era joan da kotxe barria-n (cod. 342, 30 months, 6 w / 8 m)  
Mommy work-to go is car new-in  
'Mommy went to work in the new car'
- g. Etxe-ra etorri n-aiz-enien pelota-gaz jolastu do-t (cod. 1079, 36 months, 6 w / 11 m)  
house-to come be.S1s-when ball-with play have.S1s  
'when I came home I will play with the ball'  
Intended: *etxera etor-ten naizenien...* 'when I will come home...'
- h. Amatxi-ren etxea-n ardia ikusi dut (cod. 7032, 36 months, 5 w / 8 m)  
Grandma-of house-in sheep see have.S1s  
'I have seen a sheep in grandma's house'
- i. zu hemen geratu-ko z-ara ni-rekin? (cod. 842, 36 months, 5 words, 8 morphemes)  
you here leave-FUT have.S2s me-with  
'will you stay here with me?'
- j. osaba-k zergatik ez dauka txabola Patxik bezala? (cod. 566, 42 months, 7 w / 10 m)  
uncle-Erg why Neg own.S3 cabin Patxi-Erg like  
'why does not the uncle have a cabin like Patxi has?'
- k. txikia nintz-en-ean sehaska-n egi-ten nue-n lo (cod. 7040, 36 months, 6 w / 12 m)  
small be.1s-Past-when crib-in do-IMP have.S1s-Past sleep  
'when I was that little I slept in the crib'
- l. Eskola-ko jantokia-n ema-ten di-gu-te-n ogia oso goxoa da (cod. 7049, 48 months, 8 w / 14 m)  
School-of dining room-in give-IMP have.S3pl.IO1pl bread very tasty is  
'the bread that they give us in the school meals is very tasty'
- m. gaur Amaiur ez da ikastola-ra etorri gaixorik dauelako (cod. 536, 48 months, 8 w / 10 m)  
today Amaiur Neg is school-to come sick is-because  
'today Amaiur did not come to school because he is sick'

The ANOVA analysis revealed a significant effect of age on all the scales of the BCDI-2: vocabulary [ $F_{(14, 735)} = 54.71, p < 0.001, \eta_p^2 = 0.51$ ], nominal morphology [ $F_{(14, 735)} = 37.38, p < 0.001, \eta_p^2 = 0.42$ ], verbal morphology [ $F_{(14, 735)} = 35.99, p < 0.001, \eta_p^2 = 0.41$ ], MLU3-w [ $F_{(14, 735)} = 39.24, p < 0.001, \eta_p^2 = 0.43$ ] and

## BCDI-2 (16-30 Months)

The scores on all scales of the BCDI-2 increased significantly with age, as depicted in **Figures 2, 3** (minimal-maximum scores: 0–643 in vocabulary, 0–17 in nominal morphology, 0–36 in verbal morphology, 1–10 in MLU-w and 1–16 in MLU-m). Mean and standard deviation values of BDCI-2 scales are shown in **Table 4**.

MLU3-m [ $F_{(14, 735)} = 40.20, p < 0.001, \eta_p^2 = 0.43$ ]. Age effect on each scale was large according to Cohen (1992) and Richardson (2011).

As shown in **Table 5**, correlations between vocabulary, nominal morphology, verbal morphology, MLU3-w and MLU3-m scales were strong ( $r$  range: 0.81–0.97), especially between MLU3-w and MLU3-m ( $r = 0.97$ ). Some correlation

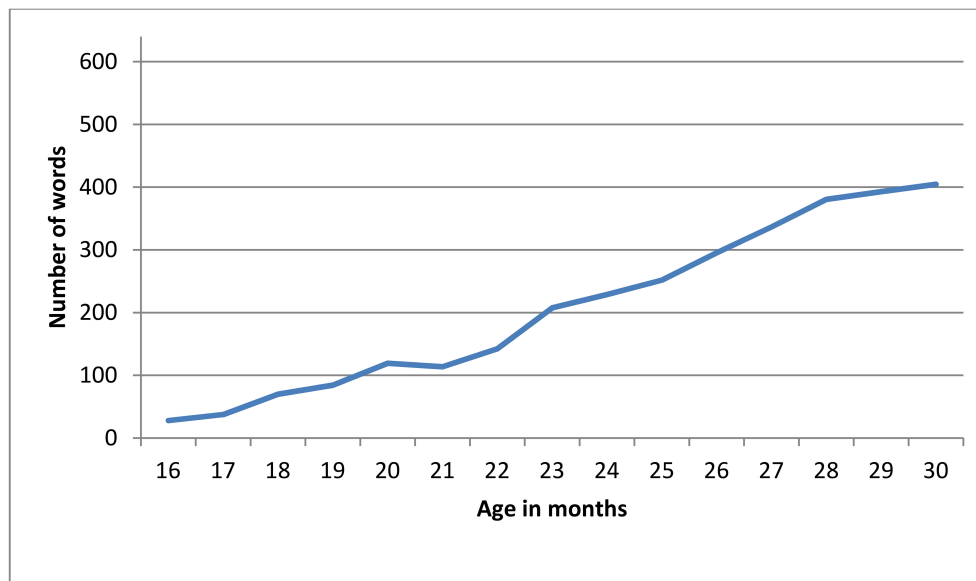


FIGURE 2 | Mean vocabulary scores by age in BCDI-2 (643 items).

coefficients decreased after controlling for age, but their values remained both significant and high ( $r$  range: 0.66–0.95). Cronbach's alpha was 0.97 for the five scales<sup>2</sup>.

### BCDI-3 (30–50 Months)

The scores on all the BCDI-3 scales increased with age, as depicted in Figures 4, 5 and the effect size of age was large. Mean and standard deviation values of BDCI-3 scales are shown in Table 6.

The ANOVA analyses revealed significant effects of age on all the BCDI-3 scales: vocabulary [ $F_{(20, 566)} = 5.46, p < 0.001, \eta_p^2 = 0.16$ ], nominal morphology [ $F_{(20, 566)} = 3.56, p < 0.001, \eta_p^2 = 0.11$ ], verbal morphology [ $F_{(20, 566)} = 5.03, p < 0.001, \eta_p^2 = 0.15$ ], on MLU3-w [ $F_{(20, 566)} = 3.822, p < 0.001, \eta_p^2 = 0.12$ ] and MLU3-m [ $F_{(20, 566)} = 4.14, p < 0.001, \eta_p^2 = 0.13$ ].

A strong correlation was found across all the BCDI-3 scales, as displayed in Table 7: vocabulary, nominal morphology, verbal morphology, MLU3-w and MLU3-m ( $r$  range: 0.55–0.97). Again, the correlation between MLU3-w and MLU3-m was particularly high ( $r = 0.97$ ). After controlling for age, some correlation coefficients decreased ( $r$  range: 0.51–0.97), but the values remained significant and high<sup>3</sup>. Cronbach's alpha was 0.91 for the five scales.

<sup>2</sup>Correlation between MLU and the scores obtained in the multiple choice question section in BCDI-2 yielded statistically significant results ( $p < 0.001$ ): MLU3-w ( $r = 0.77$  and  $r = 0.64$ , controlling for age) and MLU3-m ( $r = 0.77$  and  $r = 0.63$ , controlling for age). The analysis of the multiple choice item sections goes beyond the purpose of the current study. Nonetheless, we have reported these data because of the request of one anonymous reviewer.

<sup>3</sup>MLU and the scores obtained in the multiple choice question section in BCDI-3 also yielded statistically significant results ( $p < 0.001$ ): MLU3-w ( $r = 0.60$  and  $r = 0.54$ , controlling for age) and MLU3-m ( $r = 0.64$  and  $r = 0.59$ , controlling for age).

### Input and MLU3

Two-way ANOVA analyses were performed in order to investigate the effect of age, input (the relative amount of exposure to Basque and Spanish), and the interaction between them on the two MLU3 measures, MLU3-w and MLU3-m in the whole sample, which is depicted in Figure 6.

The first ANOVA showed main effects of both, age [ $F_{(4, 1182)} = 102.11, p < 0.001, \eta_p^2 = 0.26$ ] and input [ $F_{(3, 1182)} = 41.01, p < 0.001, \eta_p^2 = 0.09$ ] in MLU3-w and the interaction between these two variables yielded statistically significant results [ $F_{(12, 1182)} = 3.50, p < 0.001, \eta_p^2 = 0.03$ ]. See Table 8.

Further analyses on the interaction between input and age were performed by analyzing the effect of input in each age group by means of one-way ANOVAs. Regarding the analysis on MLU3-w (see Figure 6 and Table 8), no significant differences were observed across the four input groups in the youngest age group (between 18 and 24 months), [ $F_{(3, 320)} = 1.06, p = 0.364, \eta_p^2 = 0.01$ ]. However, significant differences were observed across input groups above 2 years of age: for the 25- to 30-month-olds [ $F_{(3, 367)} = 11.18, p < 0.001, \eta_p^2 = 0.08$ ], for the 31- to 36-month-olds [ $F_{(3, 165)} = 7.49, p < 0.001, \eta_p^2 = 0.12$ ], for the 37- to 42-month-olds [ $F_{(3, 175)} = 8.72, p < 0.001, \eta_p^2 = 0.13$ ] as well as for the 43- to 48-month-olds [ $F_{(3, 155)} = 10.80, p < 0.001, \eta_p^2 = 0.17$ ]. Interestingly, the size of the input effect increased with age, reaching a large size from 3 years of age (37–42 months) onwards.

Similar results were also found in MLU3-m, with significant main effects of age [ $F_{(4, 1182)} = 108.25, p < 0.001, \eta_p^2 = 0.27$ ] and input [ $F_{(3, 1182)} = 45.97, p < 0.001, \eta_p^2 = 0.10$ ]. In addition, the interaction between age and input proved significant [ $F_{(12, 1182)} = 3.99, p < 0.001, \eta_p^2 = 0.04$ ] (see Table 9).

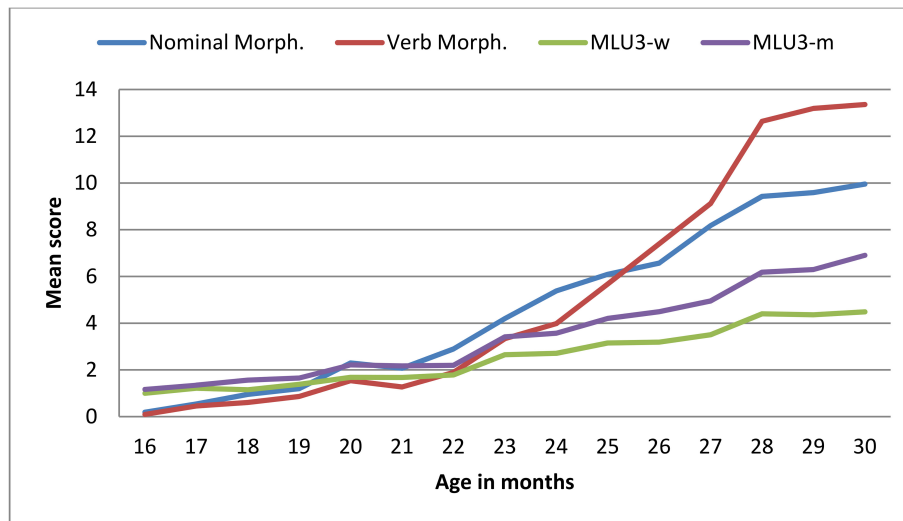


FIGURE 3 | Mean scores by age in BCDI-2 scales: nominal morphology (17 items), verbal morphology (39 items) and MLU3.

TABLE 4 | Mean scores and standard deviations of five CDI-2 scales, by age in months.

Age	N	Vocabulary		Nominal morphology		Verb morphology		MLU3-w		MLU3-m	
		M	SD	M	SD	M	SD	M	SD	M	SD
16	42	27.93	34.47	0.19	0.55	0.09	0.30	1.00	0.00	1.17	0.38
17	46	37.43	65.83	0.54	2.00	0.46	2.24	1.21	0.91	1.35	1.19
18	43	69.81	80.54	0.95	2.22	0.60	1.80	1.15	0.51	1.57	0.85
19	45	84.29	114.19	1.20	2.59	0.87	3.34	1.38	0.63	1.65	0.84
20	37	119.27	120.92	2.30	4.07	1.54	3.21	1.68	0.93	2.21	1.33
21	44	113.64	101.89	2.07	2.96	1.27	2.00	1.67	0.94	2.17	1.30
22	47	142.36	123.81	2.89	3.93	1.89	2.96	1.78	1.03	2.20	1.26
23	50	207.44	149.54	4.20	4.26	3.34	4.44	2.65	1.54	3.41	2.23
24	58	228.64	116.43	5.38	4.41	3.98	4.19	2.71	1.35	3.57	1.94
25	55	251.84	139.46	6.09	4.85	5.67	5.86	3.15	1.62	4.20	2.33
26	61	295.29	131.48	6.57	4.46	7.39	6.85	3.19	1.51	4.49	2.36
27	56	336.37	137.64	8.18	4.50	9.12	7.13	3.51	1.75	4.95	2.64
28	56	380.48	151.64	9.43	4.89	12.64	8.98	4.40	1.81	6.18	2.87
29	46	392.72	153.11	9.59	4.99	13.20	9.10	4.36	1.99	6.30	3.44
30	64	404.50	160.59	9.95	5.09	13.36	9.53	4.49	1.95	6.91	3.27
Total	750	220.52	178.68	5.02	5.24	5.51	7.52	2.69	1.84	3.70	2.86

Number of items by scale: vocabulary (643), nominal morphology (17), and verbal morphology (39).

Concerning MLU3-m (see **Figure 6** and **Table 9**), no significant differences were observed among the four input groups in the youngest age range (18–24 months) [ $F_{(3, 320)} = 1.63$ ,  $p = 0.182$ ,  $\eta_p^2 = 0.01$ ]. Nevertheless, from the age of 2 the effect of input in the MLU-w was revealed to be significant in all age groups: 25–30 months of age [ $F_{(3, 367)} = 9.73$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.07$ ], 31–36 months of age [ $F_{(3, 165)} = 7.19$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.12$ ], 37–42 months of age [ $F_{(3, 175)} = 10.37$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.15$ ] and 43–48 months of age [ $F_{(3, 155)} = 12.99$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.20$ ]. Similar to the pattern observed in MLU3-w, the size of the

input effect increased with age, reaching a large size from age 3 onwards (37–42 m).

*Post hoc* analyses with a Bonferroni correction indicated no significant differences among input groups on MLU3-w and MLU3-m scores in the youngest age group (18–24 months). However, from 2 years of age, the mean scores for monolinguals and Basque-dominant bilinguals were significantly higher than those of the Spanish-dominant bilinguals (see **Tables 8, 9**). In contrast, monolinguals and Basque-dominant bilinguals did not differ significantly throughout the whole period studied, whilst



balanced bilinguals showed intermediate scores which were closer to those of the Spanish-dominant bilingual group than to the Basque-dominant bilinguals in the age ranges before the 42nd month. Finally, in the oldest age group (43–48 months), the balanced bilinguals aligned with the Spanish-dominant bilinguals rather than with the Basque-dominant ones, as shown in **Figure 6** and **Tables 8, 9**.

Therefore, three main results can be drawn from the analyses provided above:

- 1) Large age effects were attested in MLU-w and MLU-m as well as in the rest of the scales of the BDCI-2 and BDCI-3 instruments, and high correlations were observed between both MLU scales and the other scales tested.
- 2) The two MLU scales showed almost perfect correlations.
- 3) Input groups behaved similarly in the 18–24-month-old group, but differences among input groups started to be significant from age 2 onwards, in such a way that monolingual and Basque-dominant bilinguals differed more

**TABLE 5 |** Pearson's correlations between BDCI-2 scales (and partial correlations, controlling for age).

	Nominal morphology	Verbal morphology	MLU3-w	MLU3-m
Vocabulary	0.89 (0.80)	0.83 (0.71)	0.81 (0.66)	0.82 (0.68)
Nominal morphology		0.85 (0.76)	0.82 (0.70)	0.83 (0.72)
Verbal morphology			0.82 (0.70)	0.84 (0.74)
MLU3-w				0.97 (0.95)

All correlations were significant at  $p < 0.001$ .

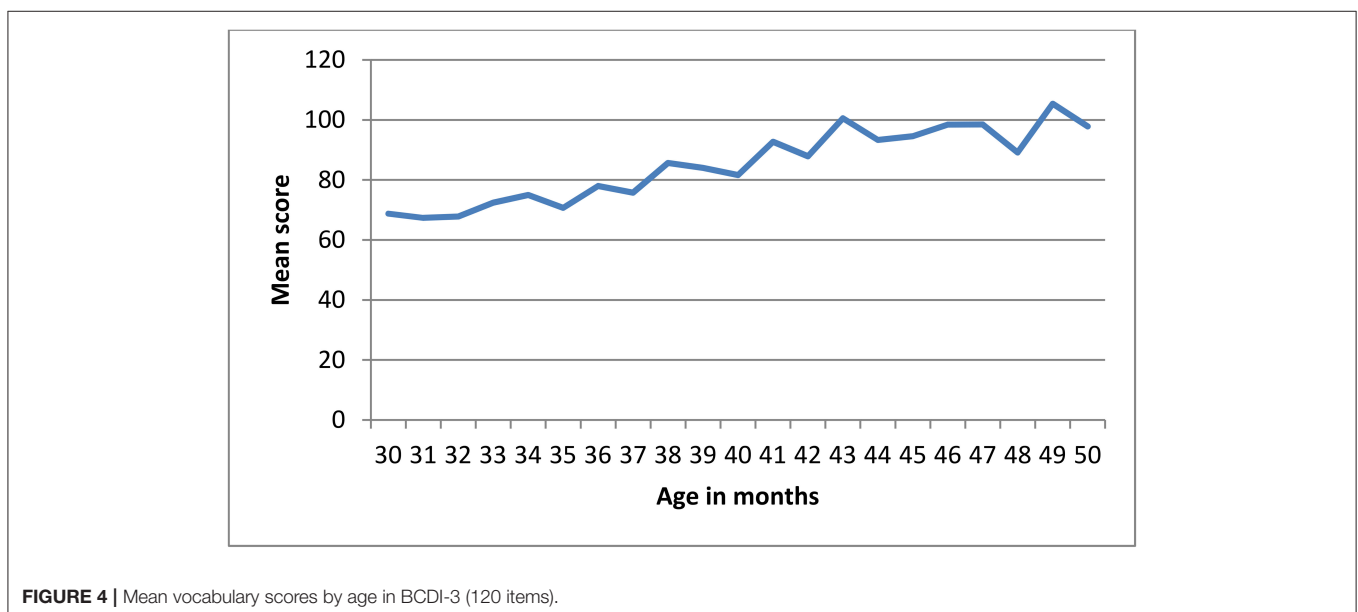
and more from the Spanish-dominant bilinguals with age, whereas the balanced bilingual group consistently showed intermediate MLU values between the groups with high (Basque-dominant) and with low (Spanish-dominant) exposure to the Basque language.

## DISCUSSION

This paper is in line with previous research which used mean length of utterance, in general, and MLU3 in particular, as an accurate index of language development for individual assessment (Brown, 1973; Fenson et al., 1993, 2007). The present bilingual data further indicate that an appropriate use of the measurement which takes into account the amount of exposure to which children are exposed will favor a more accurate assessment of these children's actual language development.

The current study, which reported MLU data of Basque obtained by means of parental questionnaires from 16- to 50-month-olds, challenged general objections regarding the reliability (a) of parental reports to assess children's expressive language, (b) of MLU as an index for language development, and (c) the accuracy of measuring MLU in words in an agglutinative language with complex morphology.

Subjectivity is one of the strongest criticisms made regarding the CDI instrument in general and the MLU3 measure in particular. Nevertheless, many studies have defended the ecological validity of parental reports as compared to studies based on experimental data, based on the observation that parents witness their children's language use in manifold communicative situations (Institute of Medicine, 2001; American Academic of Pediatrics, 2003; O'Neil, 2007). Moreover, many handbooks of the adaptations of the CDI instruments to English and many other languages include validity studies comparing CDI parental report data with data obtained using other



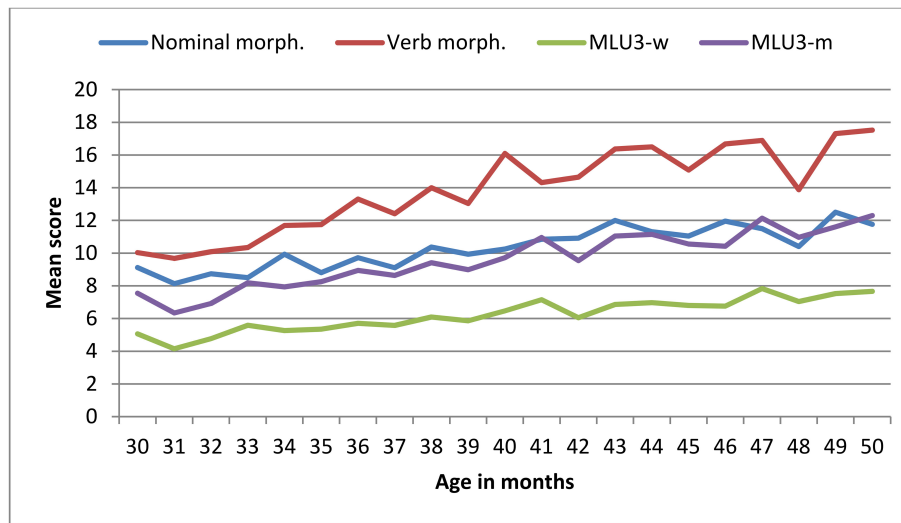


FIGURE 5 | Mean scores by age in BCDI-3 scales: nominal morphology (14 items), verbal morphology (22 items) and MLU3.

TABLE 6 | Mean scores (M) and standard deviations (SD) of five BCDI-3 scales by age in months.

Age	N	Vocabulary		Nominal morphology		Verbal morphology		MLU3-w		MLU3-m	
		M	SD	M	SD	M	SD	M	SD	M	SD
30	33	68.76	24.73	9.12	2.71	10.03	4.50	5.06	1.71	7.55	2.83
31	22	67.36	34.06	8.14	4.66	9.68	5.89	4.15	2.07	6.34	3.33
32	23	67.83	33.30	8.74	4.39	10.09	6.95	4.77	2.35	6.92	4.06
33	32	72.41	25.41	8.50	3.90	10.34	5.43	5.59	3.31	8.18	5.07
34	29	74.96	20.99	9.93	2.71	11.69	5.52	5.26	2.16	7.94	3.47
35	31	70.68	28.41	8.81	3.92	11.74	5.93	5.35	2.31	8.25	4.00
36	32	77.97	30.20	9.72	3.70	13.31	6.25	5.71	2.13	8.94	3.77
37	29	75.76	31.20	9.10	4.06	12.41	6.26	5.58	2.37	8.64	4.16
38	35	85.66	30.16	10.37	3.42	14.00	6.48	6.09	2.54	9.41	4.25
39	29	84.00	26.22	9.93	3.98	13.03	6.58	5.86	3.20	8.99	5.06
40	20	81.60	23.26	10.25	3.81	16.10	5.74	6.47	2.66	9.72	3.82
41	32	92.72	26.95	10.84	3.08	14.31	6.12	7.15	3.27	10.96	5.20
42	34	87.88	29.95	10.91	3.12	14.65	6.70	6.05	2.39	9.54	3.89
43	27	100.56	22.02	12.00	2.80	16.37	6.01	6.86	2.57	11.04	4.50
44	32	93.31	23.26	11.31	3.07	16.50	4.90	6.97	2.62	11.15	3.99
45	25	94.56	26.32	11.04	3.88	15.08	6.90	6.80	3.00	10.56	4.49
46	22	98.41	19.76	11.95	2.48	16.68	4.41	6.76	2.71	10.42	4.00
47	28	98.50	24.88	11.50	3.50	16.89	5.14	7.83	2.62	12.14	5.03
48	25	89.12	29.21	10.40	3.70	13.88	7.38	7.04	3.11	10.97	4.88
49	26	105.38	15.66	12.50	1.65	17.31	4.70	7.53	2.75	11.61	4.63
50	21	97.81	25.74	11.76	3.22	17.52	5.51	7.66	2.57	12.30	4.02
Total	587	84.69	28.67	10.29	3.61	13.79	6.32	6.19	2.74	9.56	4.49

Number of items by scale: vocabulary (120), nominal morphology (14) and verbal morphology (22).

methodologies such as elicitation, or spontaneous interaction. These studies also reported strong correlations between MLU3 and the rest of the scales (Fenson et al., 1993; Jackson-Maldonado et al., 2003; López-Ornat et al., 2005; Barreña et al., 2008a). As for the subjectivity in coding MLU in general, and MLU3

in particular, the current study was based on data coded by two different researchers for both BCDI-2 and BCDI-3 data. The high correlation found between the two analyses confirmed the reliability of the coding used. The Basque sample data of 1337 children between 16 and 50 months of age obtained with

either the BCDI-2 or the BCDI-3 revealed a gradual increase of mean scores in the scales studied throughout the age groups, month by month, similar to the one found in the lexical and grammatical scales of the BCDI-2 and BCDI-3. The high correlations found between MLU3-w, MLU3-m and the scales of vocabulary, verbal morphology, nominal morphology as well as with the section of multiple choice items regarding children's advance in the acquisition of some particular structures revealed an extremely strong internal consistency throughout the two parental questionnaires. Such a consistency proves, first, parental reports' trustworthiness when reporting about their children's language use and, second, BCDI instruments' reliability.

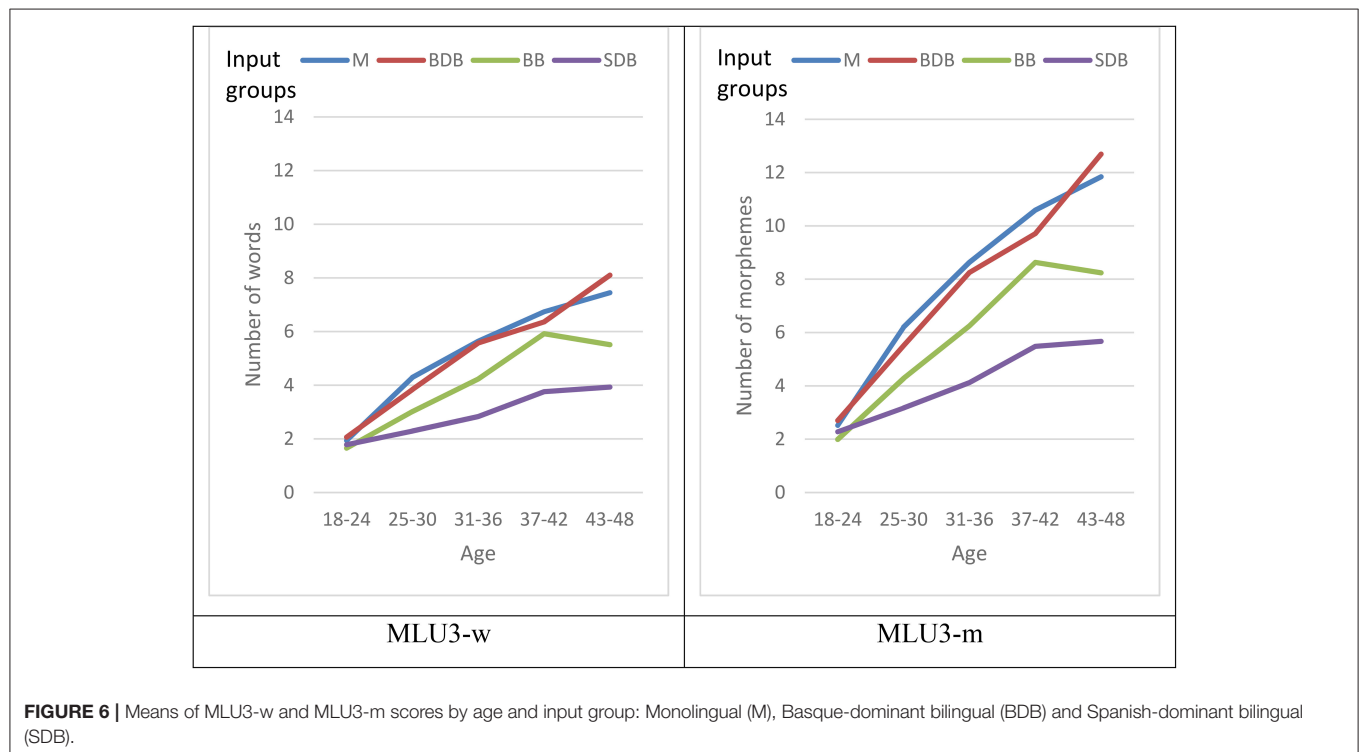
The first prediction—that MLU3 scales in BCDI would be as sensitive as the rest of the scales in this instrument in detecting

toddlers' developmental changes—has been confirmed by the data analyzed. On the one hand, the large size of age effects on the BCDI scales tested confirmed the sensitivity of MLU-w and MLU-m as well as the rest of the scales in detecting developmental changes in both instruments ( $\eta_p^2 = 0.43$  in BCDI-2,  $\eta_p^2 = 0.12$ – $0.13$  in BCDI-3). The effect size in the rest of scales was  $\eta_p^2 = 0.41$ – $0.51$  in BCDI-2 and lower, but still large or close to it ( $\eta_p^2 = 0.11$ – $0.16$ ) in BCDI-3. The fact that the effect size of age decreased from BDCI-2 ( $\eta_p^2 \approx 0.40$ ) to BCDI-3 ( $\eta_p^2 \approx 0.15$ ) can be explained in two ways. First, methodological differences such as the number of items included in the two instruments (see **Table 1**) may be the reason, at least partially, for the difference in the effect of age: the differences in the number of items are large in vocabulary (643/120 words). However, they are not so big in morphology (17/16 in nominal morphology and 40/20 in verbal morphology) where, nevertheless, the effect size of age decreased at the same pace as for vocabulary. Moreover, MLU scales were calculated in exactly the same way in both instruments and revealed again a weaker effect of age in BCDI-3 than in BDCI-2, questioning the relevance of the methodological account for the differences mentioned. The second explanation in terms of development appears to be much more convincing: the difference attested between the two Basque instruments is compatible with the stronger developmental changes taking place between the earlier developmental period covered by the BCDI-2 (16–30 months), as compared to the later one covered by the BCDI-3 (30–50 months). The decrease in developmental speed found in the Basque data is in line with that found by Fenson et al. (2007) with the English instruments CDI-2 (16–30 m) and CDI-3

**TABLE 7** | Pearson's correlations between BCDI-3 scales (and partial correlations, controlling for age).

	Nominal morphology	Verbal morphology	MLU3-w	MLU3-m
Vocabulary	0.83 (0.81)	0.76 (0.73)	0.57 (0.51)	0.60 (0.54)
Nominal morphology		0.78 (0.76)	0.55 (0.51)	0.58 (0.54)
Verbal morphology			0.58 (0.52)	0.60 (0.55)
MLU3-w				0.97 (0.97)

All correlations were significant at  $p < 0.001$ .



**TABLE 8** | Mean scores and standard deviations in MLU3-w scale across age and input groups.

Age	Input	Mean	Standard deviation	N
18–24	Monolingual	1.94	1.26	200
	Basque-dominant bilingual	2.06	1.23	63
	Balanced bilingual	1.65	0.89	41
	Spanish-dominant bilingual	1.78	1.23	20
25–30	Monolingual	4.29 a	1.75	221
	Basque-dominant bilingual	3.84 ab	1.89	96
	Balanced bilingual	3.02 bc	1.97	35
	Spanish-dominant bilingual	2.29 c	1.49	19
31–36	Monolingual	5.64 a	2.43	107
	Basque-dominant bilingual	5.57 ab	2.41	26
	Balanced bilingual	4.23 abc	1.97	22
	Spanish-dominant bilingual	2.83 c	1.95	14
37–42	Monolingual	6.74 a	2.40	109
	Basque-dominant bilingual	6.36 a	2.77	26
	Balanced bilingual	5.92 a	3.62	20
	Spanish-dominant bilingual	3.76 b	2.31	24
43–48	Monolingual	7.45 a	2.49	102
	Basque-dominant bilingual	8.10 a	2.91	26
	Balanced bilingual	5.51 b	2.52	19
	Spanish-dominant bilingual	3.93 b	1.87	12

Means that do not share a common alphabetical subscript differ at  $p < 0.05$  ( $a > b > c$ ) according to post hoc analyses with a Bonferroni correction.

(30–42 m), and with Brown's statement that MLU scales may not be accurate enough for measuring language complexity once the child has reached Stage V. Note that two of the children studied by Brown reached that stage at around age 4, whilst the third one had reached it almost 2 years earlier. Hence, this is compatible with the idea that the effect of this factor decreases after some age between 3 and 4 years.

On the other hand, the high correlations between MLU and the rest of the scales reveals the consistency of the instrument and its validity to measure children's verbal communicative development between 16 and 50 months of age in line with the results of many adaptations of the CDI-2 and CDI-3 instruments (Fenson et al., 1993, 2007; Jackson-Maldonado et al., 2003; López-Ornat et al., 2005). Even though the explanation is not clearly formulated yet, we can conclude, in line with Dethorne et al. (2005), that the strong correlation attested between MLU values and scales of varied instruments used across studies to measure children's development in different language components (expressive vocabulary, grammar...) confirms Brown's assumption that MLU is a measure of early development in language complexity in general, rather than of a specific language component, such as semantics or morphosyntax, in particular. Its validity may be limited to the earliest stages, applying no further than Stage V. Nonetheless, this last point

**TABLE 9** | Mean scores and standard deviations in MLU3-m scale across age and input groups.

Age	Input	Mean	Standard deviation	N
18–24	Monolingual	2.52	1.77	200
	Basque-dominant bilingual	2.70	1.68	63
	Balanced bilingual	1.99	1.09	41
	Spanish-dominant bilingual	2.28	1.78	20
25–30	Monolingual	6.20 a	2.91	221
	Basque-dominant bilingual	5.51 ab	3.16	96
	Balanced bilingual	4.28 b	2.88	35
	Spanish-dominant bilingual	3.17 c	1.92	19
31–36	Monolingual	8.63 a	3.97	107
	Basque-dominant bilingual	8.25 a	4.00	26
	Balanced bilingual	6.25 a	3.42	22
	Spanish-dominant bilingual	4.12 b	3.14	14
37–42	Monolingual	10.59 a	3.91	109
	Basque-dominant bilingual	9.71 a	4.33	26
	Balanced bilingual	8.63 ab	5.37	20
	Spanish-dominant bilingual	5.48 b	3.78	24
43–48	Monolingual	11.84 a	4.11	102
	Basque-dominant bilingual	12.69 a	4.44	26
	Balanced bilingual	8.24 b	3.68	19
	Spanish-dominant bilingual	5.67 b	2.58	12

Means that do not share a common alphabetical subscript differ at  $p < 0.05$  ( $a > b > c$ ) according to post hoc analyses with a Bonferroni correction.

could not be either confirmed or disconfirmed by the Basque data and requires further research.

The second hypothesis that MLU3-m would turn out to be more discriminative than MLU-w has not been confirmed by the data, since no size differences were found in the effect of age in the two MLU scales:  $\eta_p^2 = 0.43$  in BCDI-2 and  $\eta_p^2 \approx 0.11$  in BDCI-3. Moreover, the almost perfect correlations between the two MLU scales indicate their similar validity to measure utterance length, regardless of the specific unit (word/morpheme) adopted as baseline. Based on the high correlations found in studies comparing MLU-w and MLU-m scores in several languages (and even MLU counted in syllables), many authors consider that both MLU measures function equally well for measuring grammatical development (Hickey, 1991; Aguado, 1995; Parker and Brorson, 2005). In contrast, Wiczorek (2010) considers that each MLU scale measures development in a different language component: MLU-w being more related to lexical development, and MLU-m to morphological development. Our data support the former position. The high correlations between the two scales in both instruments ( $r > 0.97$  and  $r > 0.95$ , when age is controlled) confirm the utility of both indexes to measure development in language complexity. Moreover, regardless of measuring MLU3 in words or in morphemes, correlations between MLU3-m and the rest of the scales are almost identical

to those between MLU3-w and the same scales, regardless of the lexical or grammatical character of them, in contrast to what has been suggested by Wieczorek (2010). The relations across MLU measurements and between MLU3-w and MLU3-m and the rest of scales may vary across languages or language types which differ in degree of morphological complexity and transparency (agglutinative, fusionant, polysynthetic...), but such an analysis goes far beyond the scope of the current paper.

Utterance segmentation in words is much quicker and easier, since no technical descriptions are necessary, fewer decisions are required (less subjectivity) and variability across coders decreases considerably, in line with previous studies (Hickey, 1991; Jackson-Maldonado and Conboy, 2007, among others). The redundancy of using both, in addition to the ease of segmenting the utterance in words as compared to morphemes, leads us to recommend MLU-w as a more parsimonious measurement for screening in clinical studies, as has been suggested in other languages (Hickey, 1991; Parker and Brorson, 2005), without denying MLU-m's utility for more specific surveys in research.

The third hypothesis, that the relative amount of input would affect children's MLU, has been partially confirmed. MLU3 scales proved sensitive to detect input effects. A subsample of around 1200 children aged 18–48 months was analyzed with more detail in order to test MLU3's utility to test children's attained developmental level in the acquisition of a minority language in permanent contact with another socially dominant Romance language (Spanish or French). The data revealed MLU3-w and MLU3-m's sensitivity not only to age, already tested in Basque as in many other languages, but to the relative amount of exposure to the language. However, the effect of the amount of (relative) exposure to the language was not visible in the youngest child group (18–24 months). Interestingly, the effect of input increased with age after age 2, varying from medium at age 2 ( $\eta_p^2 = 0.07$  and  $0.12$ ) to large at age 3 ( $\eta_p^2 = 0.15$  and  $0.20$ ). From age 2 onwards, children with a large amount of exposure to Basque (M and BDB groups) showed more similar scores in MLU3-w and MLU3-m scales than the group with less exposure (SDB), in line with previous studies which tested these populations' lexical and grammatical scores (Barreña et al., 2008a,b).

Despite the strong intralinguistic correlations found among the BCDI subscales, in line with CDI data of English-Spanish bilinguals (Marchman et al., 2004; Hoff et al., 2014), measuring Basque bilinguals' language use only in Basque leads us to underscore the real language capacity of most participants in the present study. Children who are exposed to more than one language rarely have the same amount of exposure to one of the languages as compared to age-matched monolinguals, on whom

normative data are based (Ezeizabarrena et al., 2017). As has been shown very convincingly by Pearson et al. (1997), bilingual assessment should ideally take place in their two languages, and in this vein, the accurate evaluation of Basque-Spanish bilinguals' communicative skills should include assessing MLU in their two languages.

## CONCLUSIONS

The analysis of cross-sectional data obtained with the BCDI-2 (16–30 months) and BCDI-3 (30–50 months) of over 1200 children revealed a strong correlation between MLU3 and expressive vocabulary in both instruments, as well as between MLU3 and morphological scales. These findings confirm the consistency of the MLU measurement, as well as that of both BCDI instruments. The results also showed that MLU3-w and MLU3-m scales can report equally well on very young children's development in the Basque language up to age 4, which leads us to recommend the easier MLU-w measurement for clinical purposes. Finally, MLU3 subscales proved sensitive to input (25–48 months), which indicates the utility of these subscales to identify developmental patterns in Basque bilinguals aged 2–4.

## ETHICS STATEMENT

This study was approved by the ethics commission of the University of the Basque Country.

## AUTHOR CONTRIBUTIONS

Data-analysis: IG and M-JE. Manuscript writing and editing: M-JE and IG.

## ACKNOWLEDGMENTS

We are very thankful to all the parents, kindergartens and schools who collaborated with the research group, as well as to Andoni Barreña, Margareta Almgren, Julia Barnes, Nekane Arratibel and the rest of the members that participated in the normative study of the BCDI. We also want to acknowledge the institutions that supported this piece of research: The University of the Basque Country, The Basque Government (IT-983-16/GIC 15/129) and the Spanish Ministerio de Economía y Competitividad/FEDER (FFI2015-68589-C2-1-P). Finally, we are in debt to Maialen Iraola, the two reviewers and a native English speaker for their helpful comments. The remaining mistakes are our own.

## REFERENCES

- American Academic of Pediatrics (2003). Family-centered care and the pediatrician's role. *Pediatrics* 112, 691–696. doi: 10.1542/peds.112.3.691
- Aguado, G. (1995). *El Desarrollo del Lenguaje de 0 a 3 años*. Madrid: CEPE.
- Allen, S. E. M., and Dench, C. (2015). Calculating mean length of utterance for Eastern Canadian Inuktitut. *First Lang.* 35, 377–406. doi: 10.1177/0142723115596648

- Almgren, M. (2000). *La Adquisición del Tiempo y Aspecto Verbal en Euskara y Castellano*. Unpublished Ph.D., University of the Basque Country.
- Barreña, A. (1995). *Gramatikaren Jabekuntza-Garapena eta Haur Euskaldunak*. Bilbao: Servicio Editorial de la Universidad del País Vasco.
- Barreña, A., and Ezeizabarrena, M. J. (1999). "Acquisition bilingüe: séparation ou fusion des codes linguistiques," in *Le Bilingüisme Précoce en Bretagne, en Pays Celtiques et en Europe Atlantique*, eds F. Favereau (Rennes: Presses Universitaires de Rennes), 225–246.

- Barreña, A., Ezeizabarrena, M. J., and Garcia, I. (2008b). "Influence of the linguistic environment on the development of the lexicon and grammar of Basque bilingual children," in *A Portrait of the Young in the New Multilingual Spain*, eds C. Perez-Vidal, M. Juan-Garau, and A. Bel (Clevedon: Multilingual Matters), 86–110.
- Barreña, A., Garcia, I., Ezeizabarrena, M. J., Almgren, M., Arratibel, N., Olano, I., et al. (2008a). *MacArthur-Bates Komunikazio Garapena Neurtzeko Zerrenda. Euskarara Egokituta-Erabilitzaileentzako Gida eta Eskuliburu Teknikoa*. Bilbao: Udako Euskal Unibertsitatea.
- Basque Government (2009). *IV Mapa Sociolingüístico*. Vitoria-Gasteiz: Servicio Central de Publicaciones del Gobierno Vasco.
- Bickerton, D. (1991). "The pace of syntactic acquisition," in *Proceedings of the Annual Meetings of the Berkeley Linguistics Society*, Vol. 17, 41–52.
- Brown, R. (1973). Development of the first language in the human species. *Am. Psychol.* 28, 97–106. doi: 10.1037/h0034209
- Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155
- Dethorne, L. S., Johnson, B. W., and Loeb, J. W. (2005). A closer look at MLU: what does it really measure? *Clin. Linguist. Phon.* 19, 635–648. doi: 10.1080/02699200410001716165
- Elosegi, K. (1998). *Kasu eta Preposizioen Jabekuntza-Garapena Haur Elebidun Batengan*. Bilbao: Servicio Editorial de la Universidad del País Vasco.
- Ezeizabarrena, J. M. (1996). *Adquisición de la Morfología Verbal en Euskera y Castellano por niños Bilingües*. Bilbao: Universidad del País Vasco.
- Ezeizabarrena, M. J., Garcia, I., Almgren, M., and Barreña, A. (2017). "Diferencias individuales en el desarrollo léxico de bilingües simultáneos tempranos euskera-castellano," in *Diferencias Individuales en la Adquisición del Lenguaje*, eds C. Rojas, and V. Oropeza (México: UNAM), 195–218.
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993). *The MacArthur Communicative Development Inventories: User's guide and Technical Manual*. San Diego, CA: Singular Publishing Group.
- Fenson, L., Marchman, V. A., Thal, D., Dale, P. S., and Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories, 2nd Edn*. Baltimore, MD: Brookes Publishing.
- Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., and Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Appl. Psycholing.* 21, 95–116. doi: 10.1017/S014271640001053
- Garcia, I., Barreña, A., Ezeizabarrena, M. J., Almgren, M., Arratibel, N., and Barnes, J. (2014). Haur euskaldunen komunikazio-garapena neurtzen 30-50 hilabete bitartean: MacArthur-Bates CDI-III tresnaren euskal bertsioa. *Uztaro* 88, 33–72.
- Hickey, T. (1991). Mean length of utterance and the acquisition of Irish. *J. Child Lang.* 18, 553–569. doi: 10.1017/S0305000900011247
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., and Parra, M. (2012). Dual language exposure and early bilingual development. *J. Child Lang.* 39, 1–27. doi: 10.1017/S0305000910000759
- Hoff, E., Welsh, S., Place, S., and Ribot, K. (2014). "Properties of dual language input that shape bilingual development and properties of environments that shape dual language input," in *Input and Experience in Bilingual Development*, eds T. Grüter, and J. Paradis (Amsterdam; Philadelphia, PA: John Benjamins), 119–140.
- Idiazabal, I. (1991). Haur txikien hizkuntza heldutasuna neurtzeko baliapide bat: esaldien batzuek besteko luzera EBBL edo MLU. *Anuario del Seminario de Filología Vasca "Julio de Urquijo"*. 14, 1289–1295.
- Institute of Medicine, Committee on Quality Health Care in America (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press.
- Jackson-Maldonado, D., and Conboy, B. T. (2007). "Utterance Length Measures for Spanish-speaking Toddlers: The Morpheme versus Word Issue Revisited," in *Communication Disorders in Spanish Speakers: Theoretical, Research and Clinical Aspects*, eds J. G. Centeno, R. T. Anderson, and L. K. Obler (Clevedon; Buffalo, NY; Toronto, ON: Multilingual Matters), 142–155.
- Jackson-Maldonado, D., Thal, D. J., Fenson, L., Marchman, V., Newton, T., and Conboy, B. (2003). *MacArthur Inventarios del Desarrollo de Habilidades Comunicativas. User's Guide and Technical Manual*. Baltimore, MD; London; Sydney, NSW: Paul Brookes.
- Johnston, J. R. (2001). An alternate MLU calculation: magnitude and variability of effects. *J. Speech Lang. Hear. Res.* 44, 156–164. doi: 10.1044/1092-4388(2001/014)
- Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* 15, 155–163. doi: 10.1016/j.jcm.2016.02.012
- Kupisch, T. (2008). "Dominance, mixing and cross-linguistic influence," in *First Language Acquisition of Morphology and Syntax: Perspectives Across Languages and Learners*, eds P. Guijarro-Fuentes, M. P. Larrañaiga and J. Cibbens (Amsterdam; Philadelphia, PA: John Benjamins), 209–234.
- Larrañaiga, M. P. (2000). *Ergative Sprachen, Akkusative Sprachen. Der Erwerb des Kasus bei Bilingualen Kindern*. Frankfurt am Main: Vervuert.
- Larrañaiga, M. P., and Guijarro-Fuentes, P. (2012a). "Collecting data and measuring competence," in *Pronouns and Clitics in Early Language*, eds M. P. Larrañaiga and P. Guijarro-Fuentes (Berlin; Boston, MA: De Gruyter Mouton), 11–44.
- Larrañaiga, M. P., and Guijarro-Fuentes, P. (2012b). Clitics in L1 bilingual acquisition. *First Lang.* 32, 151–175. doi: 10.1177/0142723711403878
- López-Ornat, S., Gallego, C., Gallo, P., Karousou, S., Mariscal, S., and Martínez, M. (2005). *Inventario de Desarrollo Comunicativo MacArthur*. Madrid: TEA.
- Marchman, V., Martínez-Sussmann, C., and Dale, P. (2004). The language-specific nature of grammatical development: evidence from bilingual language learners. *Dev. Sci.* 7, 212–224. doi: 10.1111/j.1467-7687.2004.00340.x
- Meisel, J. M. (ed.) (1991). "Principles of Universal Grammar and strategies of language use: on some similarities and differences between first and second language acquisition," in *Point-Counterpoint: Universal Grammar in the Second Language*, ed L. Eubank (Amsterdam: John Benjamins), 231–276.
- Meisel, J. M. (ed.) (1994). *Bilingual First Language Acquisition: French and German Grammatical Development*. Amsterdam: John Benjamins.
- Meisel, J. M. (2011). Bilingual language acquisition and theories of diachronic change: bilingualism as cause and effect of grammatical change. *Biling. Lang. Cogn.* 14, 121–145. doi: 10.1017/S1366728910000143
- Müller, N. (1993). *Komplexe Sätze: der Erwerb von COMP und von Wortstellungsmustern bei Bilingualen Kindern (französisch/deutsch)*, Vol. 16. Tübingen: Gunter Narr Verlag.
- Müller, N., and Kupisch, T. (2003). Zum simultanen Erwerb des Deutschen und des Französischen bei (un) ausgeglichen bilingualen Kindern. *Vox Romanica* 62, 145–169.
- Newmeyer, F. J., and Preston, L. B. (eds.) (2014). *Measuring Grammatical Complexity*. Oxford: Oxford University Press.
- Newmeyer, F. J. (2017). *The Question of Linguistic Complexity: Historical Perspective*. Available online at: <https://cdn.uclouvain.be/public/Exports%20reddot/cecl/documents/Newmeyer.pdf> (Accessed July 19, 2017).
- Nice, M. (1925). Length of sentences as a criterion of a child's progress in speech. *J. Educ. Psychol.* 16, 370–379. doi: 10.1037/h0073259
- O'Neil, D. K. (2007). The language use inventory for young children: a parent-report measure of pragmatic language development for 18- to 47-month-old children. *J. Speech Lang. Hear. Res.* 50, 214–348. doi: 10.1044/1092-4388(2007/017)
- Parker, M. D., and Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (MLUw) and mean length of utterance in words (MLUw). *First Lang.* 25, 365–376. doi: 10.1177/0142723705059114
- Pearson, B. Z., Fernández, S. C., Lewedeg, G. V., and Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Appl. Psycholing.* 18, 41–58. doi: 10.1017/S0142716400009863
- Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., and Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *J. Speech Lang. Hear. Res.* 53, 333–349. doi: 10.1044/1092-4388(2009/08-0183)
- Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educ. Res. Rev.* 6, 135–147. doi: 10.1016/j.edurev.2010.12.001
- Rollins, P. R., Snow, C. E., and Willett, J. B. (1996). Predictors of MLU: Semantic and morphological developments. *First Lang.* 16, 243–259. doi: 10.1177/014272379601604705
- Schmeiser, A., Eichler, N., Arnaus, L., and Müller, N. (2016). Mélanges interpositionnels chez des bilingues franco-allemands. Est-ce vraiment

- du code-switching? *Lang. Interact. Acquis.* 7, 238–274. doi: 10.1075/lia.7.2.04sch
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *Int. J. Biling.* 15, 426–445. doi: 10.1177/1367006911403202
- Wieczorek, R. (2010). Using MLU to study early language development in English. *Psychol. Lang. Commun.* 14:59. doi: 10.2478/v10057-010-0010-9
- Yip, V., and Matthews, S. (2006). Assessing language dominance in bilingual acquisition: a case for mean length utterance differentials. *Lang. Assess. Q.* 3, 97–116. doi: 10.1207/s15434311laq0302\_2

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Ezeizabarrena and Garcia Fernandez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Backward Dependencies and *in-Situ wh*-Questions as Test Cases on How to Approach Experimental Linguistics Research That Pursues Theoretical Linguistics Questions

Leticia Pablos<sup>1,2\*</sup>, Jenny Doetjes<sup>1</sup> and Lisa L.-S. Cheng<sup>1,2</sup>

<sup>1</sup> Leiden University Centre for Linguistics, Leiden University, Leiden, Netherlands, <sup>2</sup> Leiden Institute for Brain and Cognition, Leiden University, Leiden, Netherlands

## OPEN ACCESS

### Edited by:

Ángel J. Gallego,  
Universitat Autònoma de Barcelona,  
Spain

### Reviewed by:

Ivan Ortega-Santos,  
University of Memphis, United States  
Ekaterina Chernova,  
Universitat Autònoma de Barcelona,  
Spain

### \*Correspondence:

Leticia Pablos  
l.pablos.robles@hum.leidenuniv.nl

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 31 August 2017

**Accepted:** 11 December 2017

**Published:** 11 January 2018

### Citation:

Pablos L, Doetjes J and Cheng LL-S  
(2018) Backward Dependencies and  
*in-Situ wh*-Questions as Test Cases  
on How to Approach Experimental  
Linguistics Research That Pursues  
Theoretical Linguistics Questions.  
*Front. Psychol.* 8:2237.  
doi: 10.3389/fpsyg.2017.02237

The empirical study of language is a young field in contemporary linguistics. This being the case, and following a natural development process, the field is currently at a stage where different research methods and experimental approaches are being put into question in terms of their validity. Without pretending to provide an answer with respect to the best way to conduct linguistics related experimental research, in this article we aim at examining the process that researchers follow in the design and implementation of experimental linguistics research with a goal to validate specific theoretical linguistic analyses. First, we discuss the general challenges that experimental work faces in finding a compromise between addressing theoretically relevant questions and being able to implement these questions in a specific controlled experimental paradigm. We discuss the Granularity Mismatch Problem (Poeppel and Embick, 2005) which addresses the challenges that research that is trying to bridge the representations and computations of language and their psycholinguistic/neurolinguistic evidence faces, and the basic assumptions that interdisciplinary research needs to consider due to the different conceptual granularity of the objects under study. To illustrate the practical implications of the points addressed, we compare two approaches to perform linguistic experimental research by reviewing a number of our own studies strongly grounded on theoretically informed questions. First, we show how linguistic phenomena similar at a conceptual level can be tested within the same language using measurement of event-related potentials (ERP) by discussing results from two ERP experiments on the processing of long-distance backward dependencies that involve coreference and negative polarity items respectively in Dutch. Second, we examine how the same linguistic phenomenon can be tested in different languages using reading time measures by discussing the outcome of four self-paced reading experiments on the processing of *in-situ wh*-questions in Mandarin Chinese and French. Finally, we review the implications



that our findings have for the specific theoretical linguistics questions that we originally aimed to address. We conclude with an overview of the general insights that can be gained from the role of structural hierarchy and grammatical constraints in processing and the existing limitations on the generalization of results.

**Keywords:** backward dependencies, *in-situ wh*-questions, coreference, negative polarity items, event-related potentials, self-paced reading, parsing, grammatical constraints

## INTRODUCTION

The study of language from an experimental point of view is a relatively young field in linguistics. In particular, work connected to the *parsing* or *on-line comprehension* of sentences—our area of interest in the present research—dates back to the late 60's and early 70's and has evolved from the work of various researchers who tried to put some of Chomsky's (1965) seminal ideas to test (e.g., Bever, 1970; Levelt, 1970; Kimball, 1973; Fodor et al., 1974; among others). Leaving the origins of the field aside (see Townsend and Bever, 2001; Phillips, 2013, for an overview), in this article we discuss the approach that researchers addressing topics based on strong theoretical linguistics background have taken to conduct experimental research that provides evidence for the validity of specific theoretical questions in linguistics or for the adequacy of general properties of language, such as structural hierarchy, or dependencies.

We first discuss the challenges this type of experimental approach faces in finding a balance between addressing theoretically relevant questions and being able to implement these questions in a controlled and realistic experimental paradigm. Secondly, we discuss the fact that certain theoretical questions can only be approached after building upon the evidence provided by a series of consecutive previous studies. Several researchers in the field have targeted a specific linguistic question starting from a seemingly simple paradigm in order to build upon the results and create more linguistically complex testing scenarios over thematically related follow-up experiments. Third, we illustrate through our own work two possible ways to carry out linguistic experimental research that bears heavily on linguistic theory. On the one hand, we examine linguistic phenomena that are similar at the conceptual level but different in their specific instantiations by investigating long-distance dependencies that involve either coreference of a cataphoric pronoun, or the backward interpretation of a negative polarity item in Dutch. These two linguistic phenomena have in common that the licensee always precedes its licensor and that the cue for how to identify a licensor rests upon the hierarchical structure. Specifically, we test how the expectation for the upcoming licensor might be impacted differently by linear and structural distance. For this, we discuss two experiments by Pablos et al. (2015, submitted) using event-related potentials (ERPs). On the other hand, we examine processing of a single linguistic phenomenon in unrelated languages. Specifically, we test the on-line processing of *wh-in-situ* questions in Mandarin Chinese and French. Current theoretical approaches all posit a dependency between the left periphery (e.g., in CP) and the *in-situ wh*-phrase, regardless of whether the dependency

is established through covert movement of the *wh*-phrase to the left periphery or binding of the *wh*-phrase by a question-operator (for an overview, see Cheng, 2009; Bayer and Cheng, 2017). In processing terms, the parser does not encounter an overt cue to determine the interrogative or declarative nature of the upcoming structure until the *wh*-phrase position. At the *wh*-phrase position, the parser might need to backtrack to the left periphery to establish a dependency in order to interpret the *wh*-word. In relation to this second phenomenon, we discuss four self-paced reading experiments by Pablos et al. (submitted). Throughout the presentation of these two cases, we discuss the potential cost of simplifying a theoretically-based research question so that the empirical research can still lead to a meaningful contribution to linguistic theory. In particular, in the section Studies on the neural architecture of language we discuss how the research question can evolve from its starting point to its end point so that it becomes an empirically testable question.

## Challenges for Theoretically Informed Experimental Research in Linguistics

In general, theoretical models are posited to represent the relationships, rules, constraints, etc., that relate different linguistic entities and structures. These theoretical models tend to rely mostly on evidence coming from speakers' judgment data and from corpus data. As it will be discussed in the section Studies on the neural architecture of language, there is an ongoing debate about whether the processing of language possesses mental representations that can be directly mapped to existing theoretical models (for further discussion see Phillips et al., 2011; Lewis and Phillips, 2015; Kush et al., 2017; Parker and Phillips, 2017; among others). Based on the assumption that this mapping exists, there is a growing amount of experimental work that evaluates if existing theoretical models can be corroborated and put to test.

One of the first challenges for this type of experimental approach is finding a compromise between addressing a theoretically relevant question and being able to implement the question at hand in a controlled experimental paradigm that leads to interpretable data and credible evidence. As this approach is driven by a theoretical linguistic question, the process starts by carefully thinking of an appropriate experimental setup that can target the question in the best possible way. The choice of methodology is also dependent on the theoretical question, which means that more than one method can be considered initially. There is a core difficulty about proceeding in this manner: the simplification of the linguistic paradigm linked to the research question. In this simplification process, attention has to be paid to

two things: the first is to test with limited variables in the interest of interpretable results, and the second is the permanence of the core theoretical question to the extent that is still relevant to the discussion in the field.

Consider the licensing contexts of Negative Polarity Items<sup>1</sup> (NPIs) as an example of a hypothetical testing scenario where the main research question is to find real-time or brain signatures of different NPI licensing environments. We know from existing theoretical linguistics research that NPIs can be licensed in different types of syntactic-semantic environments (e.g., conditionals, questions, comparatives, negative structures, see Giannakidou, 2011 for a full description). Thus, if there is some correspondence between the competence that speakers have of the different NPI licensing contexts and the speakers' use of this knowledge in real-time, a possible research question that we could put forth is whether these different syntactic-semantic environments yield different processing effects or whether these effects can be unified in that, if tested, they could all result into similar brain or psycholinguistic/algorithmic signatures. However, there is one constraint, namely, it is quite challenging to test all possible licensing contexts in one go. Further, if we test all possible contexts with one single experiment, we might get un-interpretable data from the fact that there are too many factors at play that are difficult to control experimentally. We therefore might break the question down into first testing only those contexts where there is an overt licenser (such as negation) that precedes the NPI. This reduces the number of factors and allows for a more uniform set of experimental stimuli, in the sense that we can at least identify the impact of an overt licenser in the processing of NPI (sentences) on-line. Once there is enough experimental evidence coming from testing environments with an overt licenser and some consensus has been reached on how NPI licensing works online (e.g., similar brain or psycholinguistic signatures are elicited), more contexts can be introduced in the experimental repertoire and in future experimental research examining the real-time signatures of NPI licensing. Nevertheless, this will only be possible when effects due to the NPI not being licensed, for example, have been robustly replicated intra-linguistically and possibly using different experimental methods. If we turn to the research on NPI processing of approximately the last 20 years, we can see that this is precisely how researchers working on this particular research question have approached this problem. Work by Shao and Neville (1998), Saddy et al. (2004), Drenhaus et al. (2005), Vasisht et al. (2008), Xiang et al. (2009), Yurchenko et al. (2013), and Parker and Phillips (2016), just to name a few, has examined the processing of NPIs by first looking at very basic paradigms where the licenser (i.e., negation) was either absent or in an inaccessible position. From all the existing research, to our knowledge, only Drenhaus et al. (2007), Steinhauer et al. (2010) and Xiang et al. (2016) examined other licensing environments

that did not require an overt licenser (i.e., *wh*-questions in Drenhaus et al., 2007; non-veridical contexts in Steinhauer et al., 2010; and emotive predicates in Xiang et al. 2016).

Furthermore, the existing studies illustrate a lack of broad cross-linguistic research in that, except for a few studies that have examined the incremental interpretation of NPI licensing in languages such as Basque (Pablos and Saddy, 2009; Pablos et al., 2011), Mandarin Chinese (Tsai et al., 2013), Dutch (Yurchenko et al., 2013), Italian (Vespignani et al., 2009), Spanish (Pablos, 2009), and Turkish (Yanilmaz and Drury, 2013), most of the existing psycholinguistic generalizations have been made based on experimental evidence coming mainly from languages such as English and German. Further, the on-line methods used vary from the use of ERPs, to eye-tracking, self-paced reading and speeded acceptability judgments, and the questions they targeted varied in nature. In all of the studies, the *resulting effect* reflects an increase of mental processing effort or an interference effect in retrieving an element from memory, but the *observable* is different depending on the method, and cannot be univocally linked to a particular neurological/psychological process (see discussion of Poeppel and Embick's, 2005, *Granularity Mismatch Problem* in the section Studies on the neural architecture of language). Therefore, only a few generalizations can be made based on the existing experimental evidence and these generalizations come mainly from research that has examined illusory licensing effects in NPI licensing contexts (see Parker and Phillips, 2016 for an overview of these effects in the psycho/neurolinguistics literature).

## Studies on the Neural Architecture of Language

One of the recurrent questions in the current psycholinguistic and neurolinguistic literature is whether researchers assume a correspondence between grammar (or our language competence system) and the parser (or our language performance system). Under the assumption of this correspondence, these two systems are able to feed each other and are part of the same cognitive system. Without such correspondence, the two systems are assumed to work separately and to abide by different rules or processes (see Lewis and Phillips, 2015 for further discussion). The research discussed here assumes that we have one cognitive system that is in charge of handling both competence and performance. What researchers working in the field of cognitive neuroscience of language have tried to address is the need to find a compromise between the theoretical assumptions that linguists take for granted and how these assumptions might be concretely realized in neurological terms (or signatures) and how they should be interpreted (see Marantz, 2005, 2013; Poeppel and Embick, 2005; Poeppel, 2012; Poeppel et al., 2012; Embick and Poeppel, 2015). Embick and Poeppel (2015, p. 358) describe one by one the challenges of how to test in an integrated way "theories of the (psycho)linguistic type that make claims about the computations and representations that constitute grammar and aspects of language use (referred to as "Computational-Representational" (CR) Theories)" in relation to "theories that study the structure and function of the brain

<sup>1</sup>Negative polarity items are items such as *anything* in English, which must appear under certain licensors, such as negation, as we can see from the comparison between (ia) and (ib):

- (i) a. John didn't buy anything.
- b. \*John bought anything.

coming from the Neurobiology of Language (NB) and that are more implementational in character.” Further, they discuss how CR-type of theories are currently more fine-grained than the current theories on how the linguistic representations and computations are realized in the brain (NB-theories). Under Poeppel and Embick’s (2005, p. 104) and Embick and Poeppel’s (2015, p. 361) view, what makes the unification of these two theories challenging is the *Granularity Mismatch Problem* (GMP), which refers to the fact that linguistic and neurolinguistic studies of language operate with objects of different “conceptual granularity.” Linguistic computation involves a number of fine-grained distinctions and explicit computational operations, whereas neuroscientific approaches involve broader conceptual distinctions. In their words, “this mismatch prevents the formulation of theoretically motivated, biologically grounded, and computationally explicit linking hypotheses that bridge neuroscience and linguistics” Poeppel and Embick (2005, p. 104) and it makes it “difficult to establish CR/NB linking hypotheses because in general the study of how the brain computes what it computes in language is at present too coarse to link up meaningfully with the distinctions made on the CR side” (Embick and Poeppel, 2015, p. 59). Adopting the view that the development of CR theory is an essential step toward understanding NB, Embick and Poeppel (2015, pp. 360–361) suggest three different ways in which CR and NB could interplay. The first is Correlational Neurolinguistics, where CR theories of language are used to investigate the NB foundations of language and in which knowledge of how the brain computes is gained by capitalizing on CR knowledge of language. This, for instance is the type of approach that works linking theoretical and psycholinguistic work have followed (see the work by Phillips and Lau, 2004; Lewis and Phillips, 2015, for example). The second way is Integrated Neurolinguistics, where Correlational Neurolinguistics plus the NB perspective provide crucial evidence that arbitrates among different CR theories. In Integrated Neurolinguistics, it is the brain data that enriches our understanding of language at the CR-level, for example. Third and last, Embick and Poeppel (2015) suggest that there is an Explanatory Neurolinguistics way where, besides Correlational and Integrated Neurolinguistics, something about NB structure or function explains why the CR theory of language involves particular computations and representations but not others.

Research over the past 10 years on the neural signatures of language has looked for experimental evidence that could show the process of how the building up of minimal units (which ranged from constituents, to minimal phrases to morphemes) occurs in the on-line computation of language, and that could show one of the basic intrinsic properties that characterizes the language faculty, namely, hierarchical structure. Within this field of work, we can distinguish three different groups of studies: (i) those that looked at whether there is hierarchy at the sentential level and whether this can be captured in terms of brain-oscillations or specific activations in syntax-semantics related brain areas (e.g., ERP studies by Luo and Poeppel, 2007; Arnal et al., 2015; Ding et al., 2015; Nelson et al., 2017; fMRI studies by Pallier et al., 2011; Brennan et al., 2012); (ii) those that examined whether a hierarchy can be found at the word level by using either fMRI or MEG methods (e.g., Fruchter

and Marantz, 2015; Fruchter et al., 2015) and (iii) those that examined the compositionality of incremental meaning using MEG methodology (e.g., Bemis and Pylkkänen, 2011; Pylkkänen et al., 2011).

The evidence coming from the first set of studies suggests that we build sentences in small constituents as we parse them incrementally and that our brain makes clear distinctions between random word lists and sentences with different constituent length, either in a more constrained (or custom made) traditional experimental setting, or in a more natural one (e.g., Brennan et al., 2012). The evidence from the second set of studies suggests that we are aware of the constituency within words in that they show differences between morphemes that hierarchically depend on the root of the word vs. those that do not. Finally, the third set of studies provides support for the construction of semantic composition starting from minimal linguistic phrases such as *red boat* and comparing them with non-compositional contexts such as a word list, e.g., *cup, boat*.

Even though the above studies have looked at different linguistic phenomena, they all seem to point to the building up of minimal linguistic units in the brain, whether we are examining minimal linguistic units at a word, phrase or sentence level. Through the use of different methods and from evidence coming from either brain oscillations or specific brain area activations, these studies have shown that there is a way to capture the representation of constituent structure in the brain. Further, all these studies have started from very simple experimental paradigms where they examined the most minimal possible linguistic interaction and they built upon their own previous results to get to robust evidence that can lead to potential generalizations about the neurobiology of language.

## Current Test Cases: Two Ways to Conduct Strongly Theoretically Informed Experimental Studies

To illustrate some of the points made above, we discuss two ways in which we approach theoretical questions in experimental terms. The first way concerns the processing of two different linguistic phenomena, coreference and negative polarity item licensing, that are conceptually similar. Both coreference and negative polarity licensing can involve long-distance backward dependencies, where the licensee or dependent element occurs linearly before its licensor (although this configuration is not necessary for any of the two phenomena). Theoretical studies treat backward dependencies the same way as forward dependencies as structural hierarchy is the only important factor rather than linear precedence. The reasoning behind both ERP experiments is to examine if the strategies employed by the parser in the online interpretation of these two types of backward dependencies are similar, despite the different nature of the relation between the dependent element and its licensor. Even though the exact nature of the dependencies is different, both dependencies are restricted by syntactic structure. In other words, in both types of dependencies, there are positions in which the licensor can occur and positions from which it is impossible to enter into a licensing relation with the licensee. The question with respect to parsing is whether these structural restrictions are

taken into account during an on-line parsing task, and whether the two types of dependencies are similar in this respect. These two types of dependencies were tested in the same language, Dutch, using the same methodology (ERPs).

The second way concerns the processing of the same linguistic phenomena, *wh-in-situ* questions, in languages with two different question formation strategies. French has both *wh*-fronting and *in-situ wh*-question strategies and Mandarin Chinese only has the *in-situ wh*-question strategy. The reasoning behind the four self-paced reading experiments we discuss is two-fold. First, as discussed above, we aim to examine the lack of an overt cue for a dependency with the left periphery (either through movement or through binding by a question-operator), and whether the encountering of the *in-situ wh*-phrase leads to backtracking in order to interpret the *in-situ wh*-phrase. Further, we examine whether the parser adopts different parsing strategies depending on whether the language only has one single *wh*-question formation strategy (e.g., only *in-situ* in Mandarin), or two strategies (as in French). If the strategies employed by the parser in the on-line interpretation of *wh-in-situ* questions in these two languages are alike, we can claim that there is a universal heuristic for interpreting *in-situ* questions in real-time. On the other hand, if the strategies differ between the two languages, we must conclude that they depend on the question formation strategies that are available to native speakers. From a theoretical point of view, it is expected that regardless of the question formation options that each language contains, *in-situ wh*-questions should be parsed similarly, namely, they need to establish dependency in the left periphery. This hypothesis considers the scenario where the grammar and the parser proceed hand-in-hand. The alternative would be an approach that shows an asymmetry between what is expected by theoretical linguistics research and what the real-time evidence shows, where the predictions for the performance side of language would be based on experience or usage-based information. If results come up differently for the two languages, it would mean that the existence of more than one question formation strategy in a language might impact the process of interpreting *in-situ wh*-questions in real-time differently. In order to address these questions, and assuming that the grammar and the parser might be unified, we tested whether *wh-in-situ* questions are processed inherently slower than their declarative counterparts when there is no prosody or context helping the online interpretation of *wh-in-situ* questions in these languages. This is the result that the theoretical approaches will predict.

## TEST CASE 1: EVENT-RELATED POTENTIAL EXPERIMENTS ON BACKWARD DEPENDENCIES IN DUTCH

### Cataphoric Pronoun Dependencies: Search for Antecedents Only in Grammatically Licit Positions

The ERP experiment in Pablos et al. (2015) examined the processing of a backward dependency involving cataphoric pronouns, i.e., pronouns that linearly precede their antecedent. The restriction of pronominal reference can be captured under

the principles of the Binding Theory (Chomsky, 1981) that indicates the configurations in which nominal elements can or cannot establish a coreferential relation. There are three Binding Principles, each of which concerns a different type of nominal element. Principle C restricts the distribution of Referential Expressions, including proper names such as *Mary*. This Binding Principle prohibits a Referential Expression (e.g., proper name) from being bound (Chomsky, 1981). We tested if the Binding Principle C constrains the on-line comprehension of pronoun-antecedent dependencies; in particular, whether antecedents are only interpreted in relation to the preceding pronoun in grammatically licit contexts (i.e., where no grammatical constraint is violated), as in the interpretation of *Mary* in relation to the cataphoric possessive pronoun *her* in (1). This scenario can be contrasted with a scenario in which establishing the antecedent-pronoun relation violates the Binding Principle C, as in (2). In such a case, the antecedent *Mary* and the pronoun *she* cannot be interpreted as referring to the same person in (2).

- (1) **Her<sub>i</sub>** sister could not drive the car in Moscow while **Mary<sub>j</sub>** was visiting.
- (2) **She<sub>i/\*j</sub>** could not drive the car in Moscow while **Mary<sub>j</sub>** was visiting.

In order to examine whether a grammatical constraint such as Binding Principle C is applied online in (2) and not in (1) at the proper name *Mary*, the well-attested *Gender Mismatch Effect* (GMME) paradigm was used (e.g., Sturt, 2003; van Gompel and Liversedge, 2003; Kazanina et al., 2007; Yoshida et al., 2014). In this paradigm, the gender mismatch effect at the antecedent position *Mary* with respect to *his* in (3) provides evidence that the parser has tried to interpret the pronoun at the antecedent position in this context. The GMME effect is observed in behavioral studies in that longer reading times in the mismatch condition in (3) than in the match condition in (1) are obtained. Conversely, when the antecedent position in (4) is compared to (2), no reading time difference is detected since *Mary* is barred as an antecedent due to Binding Principle C.

- (3) **His<sub>k</sub>** sister could not drive the car in Moscow while **Mary<sub>j</sub>** was visiting.
- (4) **He<sub>i</sub>** could not drive the car in Moscow while **Mary<sub>j</sub>** was visiting.

Previous studies have tested these specific pronoun-antecedent configurations in English and they measured reading times via different behavioral methods (i.e., self-paced reading and eye-tracking). The ERP study by Pablos et al. (2015) that we discuss here examined what the neural reflections of the GMME were<sup>2</sup> and whether the GMME could be cross-linguistically attested.

<sup>2</sup>At the time, there existed some ERP studies on the processing of forward (antecedent-)pronoun configurations (see Osterhout and Mobley, 1995; Van Berkum et al., 2007; Xu et al., 2013), but little evidence existed about how backward pronoun dependencies were processed in ERP terms. Studies on forward pronoun dependencies have resulted in the generation of a P600 at the mismatched pronoun *she* in contrast to the matched pronoun *he* in configurations such as the one in (i) from Osterhout and Mobley (1995).

(i) **The uncle** hoped that **he/she** had picked a good wine.

## Paradigm Selection and Materials' Design

Following the self-paced reading study by Kazanina et al. (2007), Pablos et al. (2015) created four different experimental conditions in Dutch to test the sensitivity of the parser to Principle C. As in (1) and (3), two “no-constraint conditions” where the pronoun could be linked to the antecedent were introduced. This is shown in the sentences in (5) and (6), which contain a possessive pronoun that either matches (*haar* - female) or mismatches (*zijn* - male) the linearly first antecedent *Suzanne*<sup>3</sup>.

- (5) **Haar**<sub>j</sub>            teamgenoten kondigden aan            dat  
*her*            *team mates*    *announced PTC*            *that*  
**Suzanne Jansen**<sub>j</sub>    zeer            hoog  
*Suzanne Jansen*    *very*            *highly*  
gewaardeerd    werd,            maar            **Edward**<sub>i</sub>    meldde  
*appreciated*    *was,*            *but*            *Edward*<sub>i</sub>    *reported*  
niet            de            exacte            waardering.  
*not*            *the*            *exact*            *rating*

‘Her teammates announced that Suzanne Jansen was highly appreciated, but Edward did not report the exact rating.’

- (6) **Zijn**<sub>i</sub>            teamgenoten kondigden aan            dat  
*his*            *team mates*    *announced PTC*            *that*  
**Suzanne Jansen**<sub>j</sub>    zeer            hoog  
*Suzanne Jansen*    *very*            *highly*  
gewaardeerd    werd,            maar            **Edward**<sub>i</sub>    meldde  
*appreciated*    *was,*            *but*            *Edward*    *reported*  
niet            de            exacte            waardering.  
*not*            *the*            *exact*            *rating*

‘His teammates announced that Suzanne Jansen was highly appreciated, but Edward did not report the exact rating.’

The other two experimental conditions were labeled as “Principle C conditions” and contained a cataphoric nominative pronoun in feminine [*zij* in (7)] or masculine [*hij* in (8)] form. Due to Principle C, these pronouns cannot corefer with the antecedent *Suzanne* in the embedded clause.

- (7) **Zij**<sub>j</sub> kondigde aan dat **Suzanne Jansen**<sub>j</sub> zeer hoog  
*she announced PTC that Suzanne Jansen very highly*  
gewaardeerd werd,  
*appreciated was*  
maar **Monika**<sub>i</sub> meldde niet de exacte waardering.  
*but Monika reported not the exact rating*  
‘She announced that Suzanne Jansen was highly appreciated, but Monika did not report the exact rating.’

<sup>3</sup>The indexing of (5) indicates the intended reading and abstracts away from the possibility that the pronoun *haar* “her” has a referent that is not mentioned in the sentence. This is also a possibility, in particular when the sentence is embedded in a context in which the referent has already been mentioned. In the experiment, the examples were given to the participants without further context. Our results, and the results of Kazanina et al. (2007) show that the parser starts an active search for a referent within the sentence after the pronoun is encountered.

- (8) **Hij**<sub>i</sub> kondigde aan dat **Suzanne Jansen**<sub>j</sub> zeer hoog  
*he announced PTC that Suzanne Jansen very highly*  
gewaardeerd werd,  
*appreciated was,*  
maar **Edward**<sub>i</sub> meldde niet de exacte waardering.  
*but Edward reported not the exact rating*  
‘He announced that Suzanne Jansen was highly appreciated, but Edward did not report the exact rating.’

## Results and Discussion

Pablos et al. (2015) found a significant ERP amplitude difference between the no-constraint conditions in (5) and (6) at the position of the name *Suzanne* in the antecedent *Suzanne Jansen*. This difference appeared as an anterior negativity over the 300–420 ms time window, where the no-constraint mismatch condition in (6) was more negative than the no-constraint match condition in (5) at the antecedent position. Furthermore, no difference was observed in the ERP waveforms between the Principle C constrained conditions in (7) and (8).

The results from this ERP experiment on Dutch backward pronoun dependencies show that the gender mismatch results in an anterior negativity and that, unlike in forward pronoun dependencies, there is no elicitation of a P600<sup>4</sup>. The anterior negativity is interpreted to be connected to failure of meeting the expectation to find an antecedent that matches in gender with the pronoun at the antecedent position. The main conclusion that one can draw from the results is that the parser is sensitive to gender mismatch effects only when they occur in grammatically licit positions. The fact that this effect is not present in the Principle C conditions means that the parser respects structural constraints when interpreting sentences in an incremental manner.

## Backward Negative Polarity Item (NPI) Dependencies: Search for Licensors Only in Grammatically Licit Positions

Similar to the cataphoric pronoun experiment discussed in the section Cataphoric pronoun dependencies, a second ERP study (Pablos et al. submitted) tested the processing of another backward dependency, a dependency involving negative polarity items. In this experiment, the Dutch negative polarity item *ook maar iets* “anything” occurs linearly before its licensor *niet* “not.” Consider first a situation where the licensor precedes the licensee as in (9a), and compare it with a context where the NPI appears linearly before the licensor, similar to the cataphoric pronoun dependency case, as in (9b) (where the NPI appears in a sentential subject). As discussed by Hoekstra (1991) and Hoeksema (2000), the subordinate clause *Dat het meisje ook maar iets geleerd heeft*

<sup>4</sup>The P600 is an Event-related component with positive polarity whose onset usually occurs around 500–600 ms and which peaks around 600 ms. Its topographical distribution is strongest over centro-parietal electrodes. It is a component that is generally elicited by syntactic violations, and its amplitude reflects the degree of difficulty of syntactic integration or reanalysis in parsing (see Osterhout and Holcomb, 1992; Hagoort et al., 1993; Kaan et al., 2000; Friederici et al., 2001, among others). The reader is referred to the discussion in Pablos et al. (2015) for the potential reasons for the lack of a P600 in the gender mismatch comparison in (5) and (6) at the antecedent.

“that the girl has learned anything” in (9b) is within the scope of the matrix negation *niet* “not,” meaning that structurally it is in a position where the NPI can be licensed by negation<sup>5</sup>. This is not the case with the negation *niet* “not” in the subordinate clause in (9c), where the NPI *ook maar iets* “anything” has scope over the negation. In this case the negation is in a position that is too low to act as a licenser of the NPI.

- (9) a. Het is **niet** waarschijnlijk dat het meisje **ook maar iets**  
*it is not probable that the girl anything*  
 geleerd heeft.  
*learned has*  
 ‘It is not probable that the girl has learned anything.’
- b. [Dat het meisje **ook maar iets** geleerd heeft] is **niet**  
*that the girl anything learned has is not*  
 waarschijnlijk.  
*probable*  
 ‘That the girl has learned anything is not probable.’
- c. \*[Dat het meisje **ook maar iets niet** geleerd heeft] is  
*that the girl anything not learned has is*  
 waarschijnlijk.  
*probable*  
 Intended: ‘That the girl has not learned anything is probable.’

The central question of this experiment was again if the parser respects grammatical constraints which would be apparent if the parser is sensitive to the hierarchical position of the licenser. The condition of “backward” NPI such as (9b) is an excellent condition to test this as we do not expect any licenser within the sentence subject, i.e., the *dat* “that”-clause, as shown in (9c). Furthermore, if we assume an incremental interpretation of the sentence in (9b), the only overt cue that the parser encounters linearly to determine that there cannot be a licenser for the NPI within the subordinate clause is the complementizer *dat* “that” and this should be enough to determine that the licenser can only occur in the main clause. The idea was that if we increase the linear distance at positions in the sentence where the parser does not expect a licenser [i.e., any position after the NPI within the *dat* “that”-clause, indicated by [A] in (10)], it should be less costly to integrate the upcoming material incrementally than if we increase the linear distance at positions in the sentence where the licenser is highly expected [i.e., any position after the main clause verb “to be,” indicated by [B] in (10)].

- (10) [Dat het meisje **ook maar iets** [A] geleerd heeft] is  
*that the girl anything learned has is*  
 [B] **niet** waarschijnlijk.  
*not probable*  
 ‘It is not probable that the girl has learned anything.’

<sup>5</sup> According to Hoekstra (1991) and Hoeksema (2000, p. 25), fronting a clause with a NPI in it yields grammatical results. Both argue that this is due to reconstruction at Logical Form, which places clauses back in their original positions. This further allows the complement-clause in (9a) to be within the scope of the matrix negator *niet* “not.” Following their account, in this study, we assume that the NPI under examination is within the scope of the matrix clause negation and thus licensed by it.

We define the processing cost following the basic assumptions of the *Dependency Locality Theory* (DLT) proposed by Gibson (1998). Gibson proposed that two types of costs could contribute to structural complexity in real-time parsing: the storage cost and the integration cost, which draw on the same pool of working memory resources. Storage costs refers to the cost of keeping an element actively stored in memory while it cannot be interpreted and while other information in the sentence is being processed. The integration cost, on the other hand, refers to the cost of integrating a syntactic prediction at the time it can be satisfied. Further, these costs are both affected by locality, which is measured in relation to the number of new discourse referents being processed<sup>6</sup>. With respect to the processing cost that we refer to when the licenser in (10) is finally parsed, we specifically refer to the integration cost, which in this sentence is connected to the integration of the NPI with the licenser at the time the prediction for the appearance of the licenser is finally met. In previous ERP studies (e.g., Fiebach et al., 2002; Phillips et al., 2005), this integration cost has been shown to elicit a P600 at the position where the syntactic prediction is met. Further, as noted in footnote 4, its amplitude has been shown to reflect the degree of difficulty of the syntactic integration at hand; therefore, one would expect that a higher integration cost will be shown in terms of differences in the amplitude of the elicited ERP component.

### Paradigm Selection and Materials’ Design

In order to test the described contrast and implement the effects of increasing the linear distance between the NPI and negation (i.e., the licenser), Pablos et al. (submitted) introduced conditions that added one to two modifiers at either A or B positions in (10). These conditions were compared at the licenser position (i.e., negation) with a control such as (9b), where no additional material was introduced. As mentioned in the section that discusses the challenges for theoretically informed experimental research, the experimental paradigm must be carefully controlled to avoid introducing differences that can affect the results: the modifiers that were included always consisted of three words each and had no possible interference in the interpretation of the NPI besides delaying the appearance of negation<sup>7</sup>. In (11a) and (11b), we reproduce examples of the experimental materials with the modifiers that were included at the A position. Again, it was expected that this contrast would not result in a high integration processing cost (in the terms we defined above) at the licenser position (i.e., negation), as the modifiers 1 and 2 occur at a structural position where negation cannot appear.

<sup>6</sup> As we discuss in section Test Case 2, Warren and Gibson (2002) showed that the nature of the referring expression that has to be integrated while a syntactic dependency or prediction is maintained can increase the syntactic complexity of the sentence. For example, indexical first and second person pronouns generate less syntactic complexity than indefinite or definite noun phrases, since the latter refer to less accessible discourse entities.

<sup>7</sup> We consider that the content of these modifiers will not disrupt the on-line interpretation of the sentence, in particular in the search for a licenser once the NPI is parsed. In this sense, even if the modifiers are different in nature from parentheticals, we follow the assumptions made by work that used parentheticals to extend the linear distance of the elements within a long-distance dependency (see Dillon et al., 2014; Parker and Phillips, 2016, for examples).

- (11) a. Dat het meisje **ook maar iets** [over dit vak]<sub>mod1</sub>  
*that the girl anything about this subject*  
 geleerd heeft is **niet** waarschijnlijk.  
*learned has is not probable*  
 ‘It is not probable that the girl has learned anything  
 about this subject.’
- b. Dat het meisje **ook maar iets** [over dit vak]<sub>mod1</sub> [op  
*that the girl anything about this subject at*  
 de universiteit]<sub>mod2</sub> geleerd heeft is **niet**  
*the university learned has is not*  
 waarschijnlijk.  
*probable*  
 ‘It is not probable that the girl has learned anything  
 about this subject in the university.’

On the other hand, in (12a) and (12b) modifiers were added to the main clause B position, which occurs adjacent to the main verb “to be.” It was expected that this contrast would result in a higher integration cost at negation due to the modifiers occurring at a structural position where negation can appear.

- (12) a. Dat het meisje **ook maar iets** geleerd heeft is  
*that the girl anything learned has is*  
 [volgens haar docent]<sub>mod1</sub>  
*according to her lecturer*  
**niet** waarschijnlijk.  
*not probable.*  
 ‘According to her lecturer, it is not probable that the girl  
 has learned anything.’
- b. Dat het meisje **ook maar iets** geleerd heeft is  
*that the girl anything learned has is*  
 [volgens haar docent]<sub>mod1</sub>  
*according to her lecturer*  
 [vanwege haar afwezigheid]<sub>mod2</sub> **niet** waarschijnlijk.  
*due to her absence not probable.*  
 ‘According to her lecturer, it is not probable that the girl  
 has learned anything due to her absence.’

Due to the fact that the NPI appears within a sentential subject clause, it is highly probable that the licenser is a negation (and not other NPI licensing environments such as conditionals, questions, etc.). Relevantly, in comparison with previous studies, the additional modifiers do not turn the test sentence into an ungrammatical continuation but rather add just extra information, avoiding effects due to grammaticality that can confound the interpretation of the results.

There are two types of potential effects that should be differentiated in the above manipulations. One is an integration cost effect from the fact that the dependency started at the NPI has decayed and retrieval of the NPI from memory when the licenser is found would be costly, and the other is a facilitation effect from the fact that negation is highly expected (and wanted) at the time the licenser is encountered. The third effect is an effect connected to the actual incremental

integration of the added modifiers and the fact that their integration also delays the appearance of the licenser (negation). Again, if the predictions we set in the section Backward Negative Polarity Item (NPI) dependencies were met, we do not expect any effect with added modifiers in the A position [as in (11a,b)], while effects are expected in the B position [as in (12a,b)]. Moreover, we expect to find an ERP component that is associated with syntactic integration costs and a difference in the amplitude of the ERP component to occur relative to the difficulty of integrating the syntactic prediction.

## Results and Discussion

Results confirm the expected contrast between the conditions in (11a) and (11b), and those in (12a) and (12b) at the negation position, when compared with their baseline condition in (9b).

The statistical analysis of the data confirmed the presence of a significant central anterior negativity in the 200–600 ms time window at the position of negation when the control sentence in (9b) was compared to conditions (12a) and (12b) at negation. When (9b) was compared to (11a) and (11b) conditions, only a lower, non-significant difference emerged. As expected, the amplitude of the negativity showed a correlation with the position and number of modifiers in the sentence with respect to the position of negation. When modifiers are introduced at the main clause following the verb *is* (i.e., position B), the amplitude of the central anterior negativity was bigger than when modifiers are introduced within the embedded clause after the NPI (i.e., position A). This shows that the parser is sensitive to structural positions in the sentence and that it considers the grammatical constraints for encoding the search for a location where a potential licenser for the NPI can occur. Furthermore, the results show that there is a different integration cost depending on the number of modifiers that are introduced at the potential licenser position.

While observable differences support the interpretation of the research question, the exact nature of the underlying process causing the ERP difference is questionable. Within the ERP literature in sentence comprehension, sustained negativities have been found for conditions that demanded a high memory load (e.g., Kluender and Kutas, 1993; King and Kutas, 1995; Friederici et al., 1996; Müller et al., 1997; Münte et al., 1998; Fiebach et al., 2002). In particular, they were found in studies that examined processing of dependencies of different lengths, where they manipulated linear distance from the start of the dependency to the closure point. These studies compared contexts of short vs. long-distance *wh*-questions (see Fiebach et al., 2002; Phillips et al., 2005) and object vs. subject relative clause contexts (King and Kutas, 1995). Furthermore, these studies carried out two types of analysis of the data. In the classic single-word ERP analysis they examined the ERPs at the beginning (i.e., *wh*-word or relativizer) and at the end of the dependencies (verb), whereas in the multiword ERP analysis of the data, they examined the ERPs elicited at each of the words of the dependency, from the beginning

(e.g., *wh*-word) to the closure of the dependency (e.g. the verb)<sup>8</sup>.

In the data from Pablos et al. (submitted), we take the beginning of the dependency to be marked by the NPI (i.e., the licensee) and the end marked by negation (i.e., the licensor). The position of negation is therefore the position where the dependency can be completed or finally integrated. It might be reasonable to think that the observed central anterior negativity marks the overall integration of the licensor for the NPI in sentences when the licensor-licensee distance is longer relative to the control. The size of the ERP amplitude is taken to reflect the level of disruption that additional material can cause in the search for a licensor. The fact that the effect correlates with the position of the intervening material (i.e., its size is relative to the position where the licensor is most likely to occur) suggests that structural conditions play a role in this process. As discussed in the section on NPI dependencies, previous studies that examined short vs. long-distance *wh*-questions (see Fiebach et al., 2002; Phillips et al., 2005) have shown the elicitation of a P600 at the verb where the dependency is completed and have interpreted it as an integration cost related to the integration of the syntactic prediction. The fact that the type of dependency we examined is of a slight different nature (i.e., on the syntax-semantics interface) might have contributed to having a different type of ERP component elicited. Again, it should be emphasized that the study by Pablos et al. (submitted) does not examine cases of licensing failure as previous researchers have done in the experimental NPI literature. Instead, it looks at grammatical instances of NPI licensing where (a) the NPI occurs linearly preceding its licensor; and (b) what is manipulated is the delay of the occurrence of the licensor at different grammatical positions. This reasoning is a bit different in spirit from previous NPI research, but it allows us to draw a parallel between the two different kinds of backward dependencies presented in the section Test Case 1 in order to answer the question of whether the parser proceeds similarly in the strategies that it adopts when proceeding in the incremental interpretation of phenomena that occur long-distance.

## General Discussion of Experiments on Test Case 1

Summarizing the main results of the ERP experiments discussed within our first test case, we first showed that gender mismatch effects in sentences containing cataphora result in anterior negativities in the 300–420 ms time-window when the gender of the antecedent mismatches that of the pronoun in no-constraint conditions. We then observed that (a) the delay in the appearance of the licensor in a structure with fronted NPIs results in a central anterior negativity in the 200–600 ms time-window at the position of negation and

(b) the difference in ERP amplitude size for the anterior negativity reflects an increased integration cost correlated with the structural position where a licensor is allowed to appear.

The common finding of these ERP experiments is that the parser respects the grammatical restrictions posited in the two configurations. In the case of coreference, the parser did not try to link the pronoun with potential antecedents in positions where the grammar (i.e., Binding Principle C) prohibits coreference, due to *c*-command, a hierarchical relation. In the case of NPI backward licensing, only modifiers added immediately before the grammatically licit licensor affect the processing of this licensor, again because the licensor position that matters is the one in which a potential licensor can have scope over the NPI, which is a necessary condition for licensing it. Even though we are not able to directly compare the elicited ERP components (since they are generated for different stimuli and their latencies and topographies do not overlap completely), these results point to the application of grammatical constraints in the on-line interpretation of the stimuli. This idea is on a par with Parker and Phillips (2016), where dependencies that consist of subject-verb agreement or reflexive-antecedents are said to deploy the same memory access mechanisms despite differing in cue weightings.

Furthermore, if we abstract away from the elicited specific ERP components, we can claim that these results yield evidence for the existence of basic hierarchical relations in parsing. These hierarchical relations are an intrinsic property of our language capacity, therefore, the results support a one-system architecture (Lewis and Phillips, 2015), where the grammar and the parser are part of the same cognitive system (as discussed in the section that has examined the neural architecture of language). Being part of the same cognitive system does not necessarily entail that the heuristics need to come in the same form in both grammar and parser, but it seems logical to assume that some of the basic properties, such as hierarchical relations, are indeed universal and shared by both. As discussed by Phillips et al. (2011) and Kush et al. (2017), one relevant property present in both the cataphora and the backward NPI licensing cases discussed within our first test case is the directionality of the dependency, where the left-hand element provides reliable information in the prospective search for an antecedent in cataphoric dependencies and for a licensor in NPI licensing dependencies.

## TEST CASE 2: EXPERIMENTS ON *WH-IN-SITU* QUESTIONS IN MANDARIN CHINESE AND FRENCH

As a second illustration of the points raised in the Introduction, in this section, we review a set of experiments where the same linguistic phenomenon is examined cross-linguistically to investigate the generalizability of parsing processes. The difference lays in the *wh*-question formation strategies available in the two tested languages.

<sup>8</sup>Notice that there is a problem inherent to the design and to the central question of our experiment and that is that we will never be able to match all the conditions closely, since they all differ in the number of words and modifiers. One potential solution would be to look at the ERP modulation of the whole sentence in a similar manner to Phillips et al. (2005), Fiebach et al. (2002), or King and Kutas (1995).



French is a language that employs two different strategies for question formation. Even though *wh-in-situ* is an option (13b), it also allows various types of structures which involve *wh*-fronting as in (13a)<sup>9</sup>:

- (13) a. **Qui** tu as vu ?  
*who you have seen*  
 ‘Who have you seen?’
- b. Tu as vu **qui** ?  
*you have seen who*  
 ‘Who have you seen?’
- c. Marie a vu Jean.  
*Marie has seen Jean*  
 ‘Marie has seen Jean.’

Whereas French has two different question formation strategies, Mandarin Chinese only has one, which we call the *in-situ wh*-question formation strategy. As shown in (14a), in this strategy the question word *shéi* ‘who’ remains in its canonical position.

- (14) a. Ni zuótiān yùjiàn le **shéi**?  
*you yesterday meet PERF who*  
 ‘Who did you meet yesterday?’
- b. Lǐsì zuótiān yùjiàn le Zhāngsān.  
*Lisi yesterday meet PERF Zhangsan*  
 ‘Lisi met Zhangsan yesterday.’

As we can see in (13) and (14), in the case of *wh-in-situ* questions, the clause type of the sentence (question or declarative) is only apparent at the point the *wh*-word is encountered [as evidenced by the comparison between (13b) and (13c) and between (14a) and (14b)]. Crucially, no distinction can be made on the surface between these two sentences by readers as they process the sentence, unless there is prosodic or contextual information available. Therefore, sentences like those in (13b) and (14a) posit an interesting question with regard to parsing covert dependencies in that, if the sentence is read and it lacks any other kind of overt cue aiding its interpretation, there are different parsing heuristics that the parser might adopt.

The syntactic literature has claimed that although *in-situ wh*-questions have no overt movement, they are formed via a covert dependency, where the *wh*-word can either relate to the left periphery (where the clause type of the sentence is flagged) via operator-variable binding, or via covert movement at Logical Form (LF; for further discussion see Huang, 1982; Cheng, 1991, 2009; Aoun and Li, 1993; Tsai, 1994; Bayer and Cheng, 2017). The

<sup>9</sup>Both (13a,b) are used in informal French only. In more formal registers, fronting is combined with subject-verb inversion or insertion of the question particle *est-ce que*. There are various pragmatic and grammatical differences between the fronting structure in (13a) and the *in-situ* one in (13b) as well as between the different possible fronting structures. For instance, the question word *pourquoi* is claimed to be bad in *in-situ* questions, while it is perfectly grammatical in most fronting questions, including the type illustrated in (13a) (Behnstedt, 1973). A full comparison between the different factors that may play a role in the choice between question strategies in French is beyond the scope of this article (but see, for instance, Boucher, 2010 for an overview).

theoretical proposals differ in the means by which the covert-dependency is formed, but they share the core assumption that there is a higher position in the structure (i.e., SpecCP) where the clause type is marked. This in turn raises an interesting question with regard to their *representation* in the language processing system. Overt dependencies have been shown to trigger active search mechanisms as soon as a fronted *wh*-word is encountered (e.g., Crain and Fodor, 1985; Stowe, 1986), but the mechanism that the parser follows in interpreting *in-situ wh*-questions is not clear since there is no trigger (or cue) for a search for a *wh*-word/phrase. Therefore, the research questions that the current test case addresses are: (a) which are the processes involved in reading *in-situ wh*-questions where no overt trigger is present for the incremental buildup of the relevant dependency? and (b) which are the observable effects of establishing the dependency in the left periphery for the *wh*-phrase?

As a first attempt we can entertain two possible approaches for the processing of *in-situ wh*-phrases: (i) the parser always posits a covert dependency from the beginning of the sentence, and therefore postulates a silent structural position at the start of the parse, or (ii) the parser only realizes it needs to establish a covert dependency when it encounters the *in-situ wh*-word/phrase. If the parser adopts the first approach, there should not be any processing cost effect observable when comparing declarative and *wh-in-situ* questions, since both are equally considered from the beginning of the parse. With the latter strategy, at the *in-situ wh*-word position, the parser will realize that a covert *wh*-dependency needs to be established, whereas this would not be necessary in declarative constructions. This effect should be similar in both Mandarin and French.

Moving one step further, it might also be possible that the integration and processing cost (see Gibson, 1998) for the covert operator position in the left periphery of a sentence differs depending on whether the language only has an *in-situ* question formation strategy (like Mandarin), or whether it is optionally *in-situ* (like French). In a language like French, once the fronted *wh*-question possibility has been discarded, the *in-situ* question continuation possibility may be less entertained. In Mandarin, where the *in-situ* strategy is the only one, the parser may anticipate the possibility of having a covert question operator, and thus encounter fewer difficulties in integrating the *in-situ wh*-expression. Thus, a further research question is: to what extent is the parser able to anticipate the upcoming structure when there is no information available to determine the *likelihood* of encountering an *in-situ* question?

The study of the processing of covert dependencies in *in-situ wh*-questions in Mandarin Chinese has already been approached in the work of Xiang et al. (2013, 2015). Xiang et al. (2013, 2015) have examined the processing of *in-situ* questions with complex *wh*-phrases with two different dependency lengths (with one embedding vs. mono clausal) and declaratives that contained definite noun phrases using different methodologies (i.e., Speed Accuracy Trade-Off (SAT), self-paced reading and eye-tracking). Xiang et al. (2013, 2015) found that *in-situ wh*-questions were processed slower, especially when *in-situ wh*-questions with one embedding were compared with mono-clausal questions. Nevertheless, there are still some

questions that remain concerning the generalizations that we can make regarding the processing of *in-situ wh*-questions. This is so because in the psycholinguistics literature both complex *wh*-phrases and definite noun phrases have been claimed to involve higher processing cost, that is, connected to the increase of the complexity of the parse, as we have discussed in the section on NPI dependencies (see also footnote 6). In complex *wh*-phrases, for example, the processing cost is said to be attributed to the discourse-linking nature of these *wh*-phrases (see De Vincenzi, 1996; Kaan et al., 2000; Donkers et al., 2013), whereas in the case of definite noun phrases, the processing cost is due to the fact that they refer to discourse entities that are less accessible and to their position in the Accessibility Hierarchy (see Ariel, 1990; Gundel et al., 1993; Warren and Gibson, 2002). Furthermore, since there is theoretical research showing that *wh*-words are closer to indefinites (see Huang, 1982; Cheng, 1991, among others), the self-paced reading experiments we report here addressed these issues connected to syntactic complexity by including an additional comparison between declarative sentences with definite and indefinite noun phrases with questions, in contexts where the *wh*-phrase was simplex (*qui* “who” and *shéi* “who”) or contexts where the *wh*-phrase was complex (such as *quel ami* “which friend” in French and *nǎgè péngyǒu* “which friend” in Mandarin Chinese).

In testing the phenomenon of *in-situ wh*-questions, Pablos et al. (submitted) wanted to compare how the incremental reading of *in-situ wh*-questions proceeds in comparison to the reading of their declarative counterparts that contain the exact same content up to the *wh*-word/noun phrase position. Their aim was two-fold: first, they wanted to investigate if the *wh*-word/phrase is expected, and if its integration is expected to be without any additional cost in comparison to its declarative counterpart; and second, they wanted to investigate whether the available *wh*-question formation strategies in each language have an impact on the initial hypotheses that are being considered by the parser before the *wh*-word/phrase position is encountered. The next section discusses the results of the four reading time experiments in Pablos et al. (submitted) on the processing of *wh-in-situ* questions in French and Mandarin Chinese.

## Processing Simplex *wh-in-Situ* Questions in French

The first of the four self-paced reading experiments in Pablos et al. (submitted) examined the contrast shown in (15) in order to test whether reading time differences can be found between questions and declaratives. To limit spurious effects, care was taken in the design of the materials: (i) the *wh*-word *qui* “who” in (15a) and the indefinite noun phrase *quelqu’un* “someone” in (15b) remain constant throughout the whole experiment; (ii) in the definite noun phrase condition, mono- and disyllabic proper names were used<sup>10</sup> to provide a match both with the length of the *wh*-word *qui* and the indefinite noun phrase *quelqu’un*, as

<sup>10</sup>Proper Names are known to result in higher processing cost than other referential noun phrases (see Warren and Gibson, 2002 for further discussion). We chose Proper Names for our design because they were the definites that consisted of single words.

illustrated in (15c); (iii) all other elements among conditions were kept minimally different.

- (15) a. *In-situ question with a simplex wh-phrase*  
 Le braqueur de banque a blessé **qui** dans sa  
*the robber of bank has hurt whom on his*  
*fuite ?*  
*escape*  
 ‘Who has the bank robber hurt on his escape?’
- b. *Declarative with indefinite object noun phrase*  
 Le braqueur de banque a blessé **quelqu’un** dans sa  
*the robber of bank has hurt someone on his*  
*fuite.*  
*escape*  
 ‘The robber of the bank has hurt someone on his escape.’
- c. *Declarative with Proper Name object*  
 Le braqueur de banque a blessé **Marie/Jean** dans sa  
*the robber of bank has hurt Marie/Jean on his*  
*fuite.*  
*escape*  
 ‘The robber of the bank has hurt Marie/Jean on his escape.’

Considering the predictions of the two possible parsing approaches described above, if only a declarative interpretation was assumed from the beginning of the sentence, the parser would need to reanalyze its initial assumption, which in turn will result in reading time differences between the declarative sentences in (15b) and (15c) and the question in (15a) at the *wh*-word/noun phrase position. Conversely, if the parser considers in parallel both possible interpretations, no reading time differences are expected between the question and the declarative conditions.

Comparison of the residual reading times of the sentences in (15) shows that there is a difference in processing times between declaratives and *in-situ* questions with a simple *wh*-phrase starting from the *wh*-word/noun phrase position. The timing of this difference depends on the type of declarative. When it contains an indefinite such as *quelqu’un* “someone” in (15b), the difference between questions and declaratives occurs as soon as the *wh*-word is encountered, whereas when it contains a proper name such as *Marie* in (18c), this difference only occurs once the proper name has been interpreted at the immediately following region [i.e., the preposition *dans* “in” within the examples in (15)].

## Processing Complex *wh-in-Situ* Questions in French

The second experiment examined the processing of questions and declaratives containing complex *wh*-phrases/noun phrases. The stimuli followed the form of the simplex *wh*-question experiment, where changes between the two experiments were only implemented at the *wh*-phrase/noun phrase position. An example of a set of materials is given in (16), with a complex *wh*-phrase *quelle caissière* “which cashier” in the *wh* condition in (16a), declaratives with an indefinite noun phrase *une caissière* “a cashier” in (16b) and declaratives with a definite noun phrase *la caissière* “the cashier” in (16c).

- (16) a. *In-situ question with a complex wh-phrase*  
 Le braqueur de banque a blessé **quelle caissière**  
*the robber of bank has hurt which cashier*  
 dans sa fuite ?  
*in his escape?*  
 ‘Which cashier has the bank robber hurt on his escape?’
- b. *Declarative with indefinite object noun phrase*  
 Le braqueur de banque a blessé **une caissière** dans  
*the robber of bank has hurt a cashier in*  
 sa fuite.  
*his escape.*  
 ‘The bank robber hurt a cashier on his escape.’
- c. *Declarative with definite object noun phrase*  
 Le braqueur de banque a blessé **la caissière** dans  
*the robber of bank has hurt the cashier in*  
 sa fuite.  
*his escape.*  
 ‘The bank robber has hurt the cashier on his escape.’

The same predictions about the two possible parsing approaches described in the case of simplex *wh*-phrases are applicable to the comparison between complex *wh*-phrases and declaratives. Our results show that *in-situ* questions with a complex *wh*-phrase such as *quelle caissière* “which cashier” in (16a) are again significantly slower to read than declaratives that contain an indefinite noun phrase such as *une caissière* “a cashier” in (16b). Interestingly, this effect is not apparent until the whole *wh*-phrase has been processed, since the effect appears at the word immediately after the *wh*-phrase (i.e., the preposition *dans* “in”). Note that *quel(le)* is the determiner of the interrogative *wh*-phrase, and the participants clearly waited until the end of the *wh*-phrase. Furthermore, given that noun phrases in French can have post-nominal modification (*quelle caissière débordée* “which overworked cashier” or *quelle caissière de supermarché* “which grocery store cashier”), the effect might be due to readers not considering the *wh*-phrases to be completed until they reached the region immediately following the noun.

### General Discussion of Experiments on Processing French *in-Situ wh*-Questions

Results from both the simplex and complex *in-situ wh*-questions in French showed that questions containing both type of *wh*-phrases are generally processed slower than their declarative counterparts, in particular those declaratives that contain indefinite noun phrases such as *quelqu’un* “someone” in (15b) and *une caissière* “a cashier” in (16b). We discuss the implications of these findings in connection to those of the self-paced readings on Mandarin Chinese in the general discussion of experiments on test case 2.

### Processing Simplex *wh-in-Situ* Questions in Mandarin Chinese

The same paradigm as in the French experiment described in the section Processing Simplex *wh-in-situ* Questions in French was used in Mandarin Chinese by Pablos et al. (submitted), contrasting *wh-in-situ* questions with declarative sentences. In

the object position, the *wh*-word *shéi* “who” was used in the *wh*-questions and *rén* “someone” (indefinite) or a proper name (definite) in declaratives.

Again, three conditions were designed to test whether reading time differences can be found between questions and declaratives in Mandarin Chinese. As in French, care was taken to minimize the differences between conditions to avoid unintentional bias of the results due to uncontrolled effects: (i) The *wh*-word *shéi* “who” in (17a) and the indefinite noun phrase *rén* “someone” in (17b) are monosyllabic and they were not changed throughout the whole experiment, whereas the proper names in (17c) were varied in having different bisyllabic proper names all throughout<sup>11</sup>; (ii) in order to make sure that the indefinite *rén* “someone” had only the indefinite interpretation available, intensional verbs were used and the perfective marker *-le* was omitted (see Cheng and Sybesma, 1999 for further discussion); (iii) the use of intensional predicates allowed for two extra regions after the *wh*-word/phrase position, which occurs usually sentence finally in Mandarin Chinese, to avoid confounds of sentence wrap-up effects at the *wh*-word position.

- (17) a. *In-situ question with a simplex phrase*  
 那个男生 想要 求 谁 解决问题?  
 Nàgè nánshēng xiǎngyào qiú **shéi** jiějué wèntí?  
*the<sup>12</sup> boy want beg who solve problem*  
 ‘Who did the boy want to beg to solve the problem?’
- b. *Declarative with indefinite object noun phrase*  
 那个男生 想要 求 人  
 Nàgè nánshēng xiǎngyào qiú **rén**  
*the boy want beg person*  
 解决问题。  
 jiějué wèntí.  
*solve problem*  
 ‘The boy wants to beg someone to solve the problem.’
- c. *Declarative with Proper Name object*  
 那个男生 想要 求 小张  
 Nàgè nánshēng xiǎngyào qiú **Xiǎozhāng**  
*the boy want beg Xiaozhang*  
 解决问题。  
 jiějué wèntí.  
*solve problem*  
 ‘The boy wants to beg Xiaozhang to solve the problem.’

The results of *in-situ* questions with a simplex *wh*-phrase in Mandarin Chinese show that *in-situ* questions with a simplex *wh*-phrase [*shéi* “who” in (17a)], were read significantly slower than their indefinite declarative counterparts [*rén* “person/someone” in (17b)] immediately after the *wh*-word, at the verb *jiejué* “to solve.” However, at the *wh*-word position *shéi* “who” in (17a), *in-situ* questions are significantly faster than their Proper Name counterparts in (17c). This slowdown effect at the proper name is attributed to two possible reasons:

<sup>11</sup>In Mandarin Chinese proper names are at least bisyllabic. This is the reason why, bisyllabic proper names are used in the study even though their corresponding interrogative and indefinite counterparts are monosyllabic.

<sup>12</sup>*Nà* “that” in Chinese is a demonstrative which is often used without a distal demonstrative reading. In such cases, its reading is more like a definite article (see Huang, 1999).

(1) proper names in the experiment materials having two morphemes/syllables while the question word *shéi* “who” only has one<sup>13</sup> and (2) the processing of proper names in Mandarin Chinese has been shown to be more costly than the processing of common nouns (see Yen, 2007).

## Processing Complex *wh-in-Situ* Questions in Mandarin Chinese

The fourth and final experiment in Mandarin Chinese from Pablos et al. (submitted) used the same paradigm as the French experiment that tested the processing of complex *wh-in-situ* questions in French, by contrasting *wh-in-situ* questions with declarative sentences. The stimuli followed the form as the simplex *wh*-question experiment described in the previous section. It only differed in content at the position of the *wh*-phrase/noun phrase: a complex *wh-in-situ* constituent [e.g., *nǎgè tóngxué* “which classmate” in (18a)] was contrasted with complex noun phrases of two types [e.g., the indefinite *yíge tóngxué* “a classmate” in (18b) and the definite *nàgè tóngxué* “the classmate” in (18c)].

- (18) a. *In-situ question with a complex phrase*  
 那个男生 想要 求 那个同学  
*Nàgè nánshēng xiǎngyào qiú nǎgè tóngxué*  
*the boy want beg which classmate*  
 解决问题?  
*jiějué wèntí?*  
*solve problem*  
 ‘Which classmate does the boy want to beg to solve the problem?’
- b. *Declarative with indefinite object noun phrase*  
 那个男生 想要 求 一个同学  
*Nàgè nánshēng xiǎngyào qiú yíge tóngxué*  
*the boy want beg a classmate*  
 解决问题.  
*jiějué wèntí.*  
*solve problem*  
 ‘The boy wants to beg a classmate to solve the problem.’
- c. *Declarative with definite object noun phrase*  
 那个男生 想要 求 那个同学  
*Nàgè nánshēng xiǎngyào qiú nàgè tóngxué*  
*the boy want beg the classmate*  
 解决问题.  
*jiějué wèntí.*  
*solve problem*  
 ‘The boy wants to beg the classmate to solve the problem.’

The results show that, when the *wh*-phrase is encountered, *in-situ* questions with a complex *wh*-phrase in Mandarin are slower at the *wh*-determiner position of the *wh*-phrase *nǎgè* “which” than their declarative counterparts containing an indefinite (i.e., *yíge* “a”). Furthermore, the slowdown carries on to the following noun region [i.e., *tóngxué* “classmate” in (18)]. At this noun, the definite declarative is still slower than the indefinite declarative.

<sup>13</sup>As discussed in footnote 11, in order to use completely natural stimuli for Mandarin, we could only use bisyllabic proper names.

Based on these results, Pablos et al. (submitted) concluded that *in-situ* questions with a complex *wh*-phrase are processed significantly slower than declaratives with an indefinite noun phrase *at the whole phrase*; while they are only processed significantly slower than declaratives with a definite noun phrase *at the noun position*. These researchers connect processing differences at the *wh*-word *nǎgè* “which,” to the discourse-link (Pesetsky, 1987; Avrutin, 2000) related cost, a well-known fact in the processing literature (see De Vincenzi, 1996; Kaan et al., 2000; Donkers et al., 2013; and for opposite claims see Frazier and Clifton, 2002; Hofmeister and Sag, 2010, among others). This means that when no prior context is given, the discourse-link feature in *nǎgè* “which” leads to similar additional processing cost as that in the definite determiner *nàgè* “the” (assuming that definites are costlier than indefinites as discussed by Warren and Gibson, 2002). In contrast, no additional processing cost is found when processing indefinite *yíge* “a” because the indefinite does not require prior discourse information.

## General Discussion of Experiments on Mandarin Chinese *in-Situ wh*-questions

The results from the processing of *in-situ* questions with a simplex and a complex *wh*-phrase in Mandarin Chinese show that, overall, both *wh*-phrase types (i.e., simplex and complex) are processed slower than *the indefinite noun phrases* within declaratives (i.e., *rén* “someone/person” and *yíge tóngxué* “a classmate”), but these effects show different timing properties depending on whether the *wh*-phrase is complex or simplex.

Based on the hypotheses put forth in the section Test Case 2 for *wh*-question formation strategies across languages, the results obtained by Pablos et al. (submitted) for the processing of *in-situ* questions containing complex and simplex *wh*-phrases in Mandarin support the approach in which the question interpretation is only considered when the *wh*-phrase is encountered, and not before. Nevertheless, this prediction seems to only be met when differences between *in-situ wh*-questions and declaratives containing indefinite noun phrases are taken into consideration. Declaratives that contain definite noun phrases do not seem to pattern accordingly. Researchers have previously identified the reading time cost of proper names and definite noun phrases over indefinite noun phrases in the processing literature (see Warren and Gibson, 2002; Yen, 2007). Thus, this result is consistent with previous findings.

## General Discussion of Experiments on Test Case 2

In the four self-paced reading experiments on the processing of *in-situ* simplex and complex *wh*-questions in French and Mandarin Chinese, results show that both simplex and complex *wh*-questions are generally processed slower than declaratives with indefinite noun phrases. Overall, the results suggest that, as hypothesized by one of the processing strategies discussed in Test Case 2, speakers of French and Mandarin do not seem to consider the *in-situ wh*-question interpretation of the sentences until they encounter the *wh*-word/phrase. This seems to occur regardless of whether the language has different *wh*-question

formation strategies or whether the only available strategy is the *in-situ wh*-question formation.

This suggests that the same processing mechanism is used in these two languages when no prosodic or contextual information is being considered. Furthermore, the results are compatible with the theoretical analyses of *in-situ wh*-questions involving covert dependencies between the *in-situ* item and the left-periphery.

As seen in the previous sections on the Mandarin and French experiments, we matched the experimental paradigms that we used for French and Mandarin as closely as possible bearing in mind the differences between the two languages. This strong parallelism provided us with the opportunity to see which effects were maintained across languages despite their differences and which effects could connect to the restrictions imposed by the research question that we pursued and the experimental technique we used. For example, the timing and length of the observed effects did not always coincide for both languages. This is very likely to be dependent on specific characteristics of the language and the data used, which point to several processes occurring at same time (e.g., dependency completion, referential assignment, etc.). The measurement of the effects by means of reading time differences can therefore not be conclusively associated to a single processing task, but might be connected to several other processes involved in the on-line comprehension of these constructions. Nevertheless, if we consider the overall result, the observable differences confirm the presence of on-line incremental interpretational processes in both languages. The results suggest that in both languages, the parser does not postulate the possibility of a question operator in CP before encountering the *in-situ wh* expression. Furthermore, the evidence coming from a close comparison of the two languages points to the existence of a common processing strategy adopted by their speakers.

## GENERAL DISCUSSION

In the previous sections, we have discussed two ways to conduct strongly theoretically informed experimental studies. In the first test case, we examined the processing of backward dependencies using two different linguistic phenomena (the referential interpretation of cataphoric pronouns and NPI licensing), with one method and one language. In the second test case, we examined the processing of one linguistic phenomenon (*in-situ wh*-questions) in different languages using a uniform method of testing and as closely as possible matched linguistic paradigms. The objective of these two tests cases was twofold: (1) to assess whether we can find common strategies in the processing of different backward dependencies and (2) to investigate whether there is a common strategy in how *wh-in-situ* questions are processed across languages.

Considering the evidence provided by the test cases discussed within this article, we can draw two major conclusions: (1) that the parser respects grammatical constraints, which means it is sensitive to differences in (hierarchical) structure, and (2) that there is a common parsing procedure for *in-situ wh*-question parsing phenomena in languages with different question

formation strategies, where the analysis of the sentence as a *wh*-question does not seem to be assumed until overt evidence such as the *wh*-word/phrase is found in the input.

Based on what we have discussed so far, the question that remains is how our experimental results can feed theoretical linguistics or what insight can we gain from these results. In other words, how can our results contribute to the linking hypothesis discussed by Embick and Poeppel (2015). There are two possible reasons why this research can be relevant for theoretical linguistics. The first is more straightforward, as it is connected to testing the same phenomenon in different languages with different question formation options. If the existing question formation strategies in these languages do not seem to make any difference for their parsing, then it means that at some level they share some basic properties. The main syntactic analyses of *in-situ wh*-questions assume a covert dependency (either through covert movement or a question operator binding with the *in-situ* element). The reported results are consistent with the establishing of a covert dependency (without choosing the particular type of ways to establish the covert dependency). The second is a more challenging one, since it comes from phenomena that are conceptually the same but different in their realization. The argument here would be that, if we find that the parser responds similarly to hierarchical relations, despite differences in the configuration of each tested structure, then it has to be the case that the parser can extract general grammatical properties out of specific input and that it can deduce the structural hierarchy behind the linearly presented input.

As discussed in the discussion of the challenges for theoretically informed experimental research in linguistics, there is usually some simplification of the theoretical question when searching for a suitable experimental paradigm. In our test cases, the starting theoretical question is much more complex than the evidence that we obtain, which supports there being hierarchical relations, for example. This means that, as researchers, we have to be aware of there being some theoretical questions that we are not going to be able to address yet. In particular, when we consider the relative maturity of the field of experimental linguistics, our current insight on experimental methods and procedure, there still exists a margin between the pursued theoretical question and the obtained results, i.e., the so-called *Granularity Mismatch Problem* in Poeppel and Embick's (2005) terms.

Finally, on the empirical side, our results are in line with current research that is connected to strongly theoretically based questions, such as the processing of Strong and Weak Crossover dependencies. For example, the research by Kush et al. (2017) also tries to examine how an incremental parser might interpret dependencies that can only be made licit once the right-hand of the sentence is known, which is comparable to the experiments on the processing of *wh-in-situ* questions. This is crucial when we compare this type of dependencies with the backward dependency cases, where the expectation for a licensor is turned into a forward search. This implies that backward and forward processes engage different parsing processes: in the case of backward dependencies there is a search for the licensor started at the licensee (the pronoun or NPI in our test case 1), whereas in the *in-situ* questions there is a retrieval or backward search for a

licensor started at the licensee (the *wh*-word/phrase). There is an overall tendency in the field of psycholinguistics to compare the processing of dependencies with similar characteristics in terms of retrieval and attraction processes in order to shed further light into how closely the parser follows the constraints of grammar. Work from Parker and Phillips (2016, 2017), for example, has compared licensor-NPI, reflexive-antecedent and subject-verb agreement dependencies in an attempt to investigate how much these dependencies look alike in their parsing routines. Our first test case on the processing of backward dependencies connects with this research in that dependencies that seem apparently quite different in their realization can show a similar processing behavior.

To conclude, it seems to us that the only way to reach some maturity in the field of experimental linguistics research is to generate a big pool of evidence that builds upon showing some of the basic properties of language in performance across different languages, so that, with time, it will be possible to find evidence for more complex relations, enabling us to bring theory and experimental evidence closer.

## AUTHOR CONTRIBUTIONS

LP, JD, and LC conceived the project, were involved in all aspects of the design of the proposed methodology as well as on the interpretation of the results. LP was involved in the experiment creation and implementation, data analysis and contributed to drafting the manuscript. JD and LC critically revised the manuscript. All authors are responsible for final approval of the version to be published.

## ACKNOWLEDGEMENTS

We would like to thank the following colleagues for their help in different stages of this work. Yang Yang for her help in the

generation, testing, analysis, and interpretation of the self-paced reading studies in Mandarin Chinese. Bobby Ruijgrok for his help in the generation of the Dutch GMM stimuli and in testing the ERP study at Leiden University. Erik Schoorlemmer for his help in the generation of the Dutch NPI stimuli. Sylvie Cuchet, Lucas Tual, Juliette Angot, and Hamida Demirdache for their help in the testing of the self-paced reading studies in French. Xiaolu Yang for her help in conducting the self-paced readings in Mandarin Chinese. Nina Kazanina for her willingness to share her English materials for the Dutch GMM ERP study and her valuable feedback regarding the interpretation of both ERP studies. Masaya Yoshida, Ming Xiang, and Nayoung Kwon for their feedback on the analysis and interpretation of the self-paced reading studies on Mandarin Chinese. Niels O. Schiller for his feedback on the generation and interpretation of the Dutch ERP studies and for providing insightful feedback for all the studies included within this article. And last but not least, Stella Gryllia and Aliza Glasbergen-Plas for helping out tremendously with their feedback at different stages of the generation, testing, analysis, and interpretation of the experiments on *in-situ wh*-questions in Mandarin Chinese and French. Earlier versions of this work were presented at the following conferences: Architectures and Mechanism of Language Processing (AMLaP) in 2011, 2015, and 2016, CUNY Human Sentence Processing Conference in 2012, GLOW35 in 2012, CNS in 2014, the 3rd East Asian Psycholinguistics Colloquium (EAPC3) in 2015, the XII International Symposium of Psycholinguistic in 2015, and at the Language and Cognition Group (LACG) at Leiden University Centre for Linguistics (LUCL). We thank the audiences of all these conferences and discussion group for their input. The research in this article was partly funded by *The Netherlands Organisation for Scientific Research* (NWO) 10.13039/501100003246 within the project *Understanding Questions* (Grant: 360-70-480).

## REFERENCES

- Aoun, J., and Li, A. Y.-H. (1993). *Wh*-elements *in-situ*: syntax or LF? *Ling. Inquiry* 24, 199–238.
- Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London: Routledge.
- Arnal, L. H., Poeppel, D., and Giraud, A. L. (2015). Temporal coding in the auditory cortex. *Handb. Clin. Neurol.* 129, 85–98. doi: 10.1016/B978-0-444-62630-1.00005-6
- Avrutin, S. (2000). Comprehension of discourse-linked and non-discourse-linked questions by children and *Broca's* aphasics. *Lang. Brain Represent. Proces.* 295–313. doi: 10.1016/B978-012304260-6/50017-7
- Bayer, J., and Cheng, L. L. S. (2017). “*Wh-in-situ*,” in *The Blackwell Companion to Syntax, 2nd Edn.*, eds M. Everaert and H. van Riemsdijk (Hoboken, NJ: Wiley-Blackwell Publishing), 1–44. doi: 10.1002/9781118358733.wbsyncom106
- Behnstedt, P. (1973). *Viens-tu? Est-ce que tu viens? Tu viens? Formen und Strukturen des direkten Fragesatzes im Französischen (Tübinger Beiträge zur Linguistik 41)*. Tübingen: Gunter Narr.
- Bemis, D. K., and Pylkkänen, L. (2011). Simple Composition: a magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *J. Neurosci.* 31, 2801–2814. doi: 10.1523/JNEUROSCI.5003-10.2011
- Bever, T. G. (1970). “The cognitive basis for linguistic structures,” in *Cognition and the Development of Language*, ed J. R. Hayes (New York, NY: Wiley), 279–362.
- Boucher, P. (2010). “*Wh*-questions in French and English,” in *Comparative and Contrastive Studies of Information Structure*, eds C. Breul and E. Göbbel (Amsterdam: John Benjamins Publishing Company), 101–138.
- Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., and Pylkkänen, L. (2012). Syntactic complexity predicts anterior temporal activity during natural story listening. *Brain Lang.* 120, 163–173. doi: 10.1016/j.bandl.2010.04.002
- Cheng, L. L. S. (2009). *Wh-in-situ* from the 1980s to now. *Lang. Linguist. Compass* 3, 767–791. doi: 10.1111/j.1749-818X.2009.00133.x
- Cheng, L. L.-S. (1991). *On the Typology of Wh-questions*. Ph.D. dissertation, MIT, Garland Publishers.
- Cheng, L. L.-S., and Sybesma, R. P. E. (1999). Bare and not so bare nouns and the structure of NP. *Ling. Inquiry* 30, 509–542. doi: 10.1162/002438999554192
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- Crain, S., and Fodor, J. D. (1985). “How can grammars help parsers?” in *Natural Language Parsing. Psychological, Computational, and Theoretical Perspectives*, eds D. Dowty, L. Karttunen, and A. M. Zwicky (Cambridge: Cambridge University Press), 94–128.
- De Vincenzi, M. (1996). Syntactic analysis in sentence comprehension: effects of dependency types and grammatical constraints. *J. Psycholinguist. Res.* 25, 117–133. doi: 10.1007/BF01708422

- Dillon, B., Clifton, C. Jr., and Frazier, L. (2014). Pushed aside: parentheticals, memory, and processing. *Lang. Cogn. Neurosci.* 29, 483–498. doi: 10.1080/01690965.2013.866684
- Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164. doi: 10.1038/nn.4186
- Donkers, J., Hoeks, J. C. J., and Stowe, L. A. (2013). D-linking or set-restriction? Processing Which-questions in Dutch. *Lang. Cogn. Process.* 28, 9–28. doi: 10.1080/01690965.2011.566343
- Drenhaus, H., Blaszczak, J., and Schütte, J. (2007). “Some psycholinguistic comments on NPI licensing,” in *Proceedings of Sinn und Bedeutung*, Vol. 11, ed E. Puig-Waldmüller (Barcelona: Universitat Pompeu Fabra), 180–193.
- Drenhaus, H., Saddy, D., and Frisch, S. (2005). “Processing negative polarity items: when negation comes through the backdoor,” in *Linguistic Evidence-Empirical, Theoretical, and Computational Perspectives*, eds S. Kepsner and M. Reis (Berlin: Mouton de Gruyter), 145–165.
- Embick, D., and Poeppel, D. (2015). Towards a computational(ist) neurobiology of language: correlational, integrated and explanatory neurolinguistics. *Lang. Cogn. Neurosci.* 30, 357–366. doi: 10.1080/23273798.2014.980750
- Fiebach, C. J., Schlesewsky, M., and Friederici, A. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German WH-questions. *J. Mem. Lang.* 47, 250–272. doi: 10.1016/S0749-596X(02)00004-9
- Fodor, J. A., Bever, T. G., and Garrett, M. (1974). *The Psychology of Language*. New York, NY: McGraw Hill.
- Frazier, L., and Clifton, C. Jr. (2002). Processing “d-linked” phrases. *J. Psycholinguist. Res.* 31, 633–660. doi: 10.1023/A:1021269122049
- Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., and Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: a spatio-temporal analysis of event-related brain potentials. *Cogn. Brain Res.* 11, 305–323. doi: 10.1016/S0926-6410(00)00065-3
- Friederici, A. D., Mecklinger, A., and Hahne, A. (1996). The temporal structure of syntactic parsing: early and late event-related brain potential effects elicited by syntactic anomalies. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1219–1248. doi: 10.1037/0278-7393.22.5.1219
- Fruchter, J., and Marantz, A. (2015). Decomposition, Lookup, and Recombination: MEG evidence for the Full Decomposition model of complex visual word recognition. *Brain Lang.* 143, 81–96. doi: 10.1016/j.bandl.2015.03.001
- Fruchter, J., Linzen, T., Westerlund, M., and Marantz, A. (2015). Lexical preactivation in basic linguistic phrases. *J. Cogn. Neurosci.*, 27, 1912–1935. doi: 10.1162/jocn\_a\_00822
- Giannakidou, A. (2011). “Negative and positive polarity items,” in *Semantics: An International Handbook of Natural Language Meaning de Gruyter*, eds H. von Klaus, C. Maienborn, and P. Portner (Berlin: Mouton de Gruyter), 1660–1712.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1–76. doi: 10.1016/S0010-0277(98)00034-1
- Gundel, J., Hedberg, H., and Zacharski, R. (1993). Referring expressions in discourse. *Language* 69, 274–307. doi: 10.2307/416535
- Hagoort, P., Brown, C., and Groothusen, J. (1993). The syntactic positive shift as an ERP-measure of syntactic processing. *Lang. Cogn. Process.* 8, 439–483. doi: 10.1080/01690969308407585
- Hoeksema, J. (2000). “Negative polarity Items: triggering, scope and c-command,” in *Negation and Polarity: Syntactic and Semantic Perspectives*, eds L. R. Horn and Y. Kato (Oxford; New York, NY: Oxford University Press), 147–192.
- Hoekstra, E. (1991). *Licensing Conditions on Phrase Structure*. Ph.D. dissertation. University of Groningen.
- Hofmeister, P., and Sag, I. A. (2010). Cognitive constraints and island effects. *Language (Baltim)* 86, 366–415. doi: 10.1353/lan.0.0223
- Huang, C. -T. J. (1982). *Logical Relations in Chinese and the Theory of Grammar*. Ph.D. Dissertation, MIT.
- Huang, S. F. (1999). The emergence of a grammatical category definite article in spoken Mandarin. *J. Pragmat.* 31, 77–94. doi: 10.1016/S0378-2166(98)00052-6
- Kaan, E., Harris, A., Gibson, E., and Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Lang. Cogn. Process.* 15, 159–201. doi: 10.1080/016909600386084
- Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., and Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *J. Mem. Lang.* 56, 384–409. doi: 10.1016/j.jml.2006.09.003
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition* 2, 15–47. doi: 10.1016/0010-0277(72)90028-5
- King, J. W., and Kutas, M. (1995). Who did what and when? using word- and clause-level ERPs to monitor working memory usage in reading. *J. Cogn. Neurosci.* 7, 376–395. doi: 10.1162/jocn.1995.7.3.376
- Kluender, R., and Kutas, M. (1993). Bridging the gap: evidence from ERPs on the processing of unbounded dependencies. *J. Cogn. Neurosci.* 5, 196–214. doi: 10.1162/jocn.1993.5.2.196
- Kush, D., Lidz, J., and Phillips, C. (2017). Looking forwards and backwards: the real-time processing of Strong and Weak Crossover. *Glossa* 2, 1–29. doi: 10.5334/gjgl.280
- Levelt, W. J. M. (1970). “A scaling approach to the study of syntactic relations,” in *Advances in Psycholinguistics*, eds G. B. Flores d’Arcais and W. J. M. Levelt (New York, NY: American Elsevier), 109–121.
- Lewis, S., and Phillips, C. (2015). Aligning grammatical theories and language processing models. *J. Psycholinguist. Res.* 44, 27–46. doi: 10.1007/s10936-014-9329-z
- Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004
- Marantz, A. (2005). Generative linguistics within the cognitive neuroscience of language. *Ling. Rev.* 22, 429–445. doi: 10.1515/tlir.2005.22.2-4.429
- Marantz, A. (2013). No escape from morphemes in morphological processing. *Lang. Cogn. Process.* 28, 905–916. doi: 10.1080/01690965.2013.779385
- Müller, H. M., King, J. W., and Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cogn. Brain Res.* 5, 193–203. doi: 10.1016/S0926-6410(96)00070-5
- Münste, F., Schiltz, K., and Kutas, M. (1998). When temporal terms belie conceptual order. *Nature* 395, 71–73. doi: 10.1038/25731
- Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., et al. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proc. Natl. Acad. Sci. U.S.A.* 114, E3669–E3678. doi: 10.1073/pnas.1701590114
- Osterhout, L., and Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806. doi: 10.1016/0749-596X(92)90039-Z
- Osterhout, L., and Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *J. Mem. Lang.* 34, 739–773. doi: 10.1006/jmla.1995.1033
- Pablos, L. (2009). “Subjunctive mood and NPI licensing in Spanish,” in *Poster Presented at Architectures and Mechanisms for Language Processing AMLaP 2009* (Barcelona: University Pompeu Fabra).
- Pablos, L., and Saddy, J. D. (2009). “NPIs and complementizer agreement in basque,” in *Papers from Brain Talk: Discourse with and in the Brain*, eds K. Alter, M. Horne, M. Lindgren, M. Roll, and J. von Koss Torkildsen (Lund: Lund University), 3–21.
- Pablos, L., Doetjes, J., Ruijgrok, B., and Cheng, L. L.-S. (2015). Active search for antecedents in cataphoric pronoun resolution. *Front. Psychol.* 6:1638. doi: 10.3389/fpsyg.2015.01638
- Pablos, L., Shirley, E., Erdocia, K., Laka, I., Williams, N., and Saddy, J. D. (2011). “Licensing of NPIs in Basque: an ERP Study,” in *Poster Presented at the 10th International Symposium of Psycholinguistics* (Donostia).
- Pallier, C., Devauchelle, A., and Dehaene, S. (2011). Cortical Representation of the Constituent Structure of Sentences. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2522–2527. doi: 10.1073/pnas.1018711108
- Parker, D., and Phillips, C. (2016). Negative polarity illusions and the format of hierarchica encodings in memory. *Cognition* 157, 321–339. doi: 10.1016/j.cognition.2016.08.016
- Parker, D., and Phillips, C. (2017). Reflexive attraction in comprehension is selective. *J. Mem. Lang.* 94, 272–290. doi: 10.1016/j.jml.2017.01.002
- Pesetsky, D. (1987). “Wh-in-situ: movement and unselective binding,” in *The Representation of (In)definiteness*, eds E. J. Reuland and A. ter Meulen (Cambridge, MA: MIT Press), 98–129.
- Phillips, C. (2013). “Parser-grammar relations: we don’t understand everything twice,” in *Language Down the Garden Path: The Cognitive Basis for Linguistic Structure*, eds M. Sanz, I. Laka, and M. Tanenhaus (Oxford: Oxford University Press), 294–315.

- Phillips, C., and Lau, E. (2004). Foundational issues. *J. Linguist.* 40, 571–591. doi: 10.1017/S002226704002774
- Phillips, C., Kazanina, N., and Abada, S. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cogn. Brain Res.* 22, 407–428. doi: 10.1016/j.cogbrainres.2004.09.012
- Phillips, C., Wagers, M. W., and Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. *Exp. Inter. Syntax Semant.* 37, 148–180. doi: 10.1108/S0092-4563(2011)0000037009
- Poeppl, D. (2012). The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Cogn. Neuropsychol.* 29, 34–55. doi: 10.1080/02643294.2012.710600
- Poeppl, D., and Embick, D. (2005). “The relation between linguistics and neuroscience,” in *Twenty-First Century Psycholinguistics: Four Cornerstones*. ed A. Cutler (Mahwah, NJ: Lawrence Erlbaum Associates), 103–120.
- Poeppl, D., Emmorey, K., Hickok, G., and Pytkäinen, L. (2012). Towards a new neurobiology of language. *J. Neurosci.* 32, 14125–14131. doi: 10.1523/JNEUROSCI.3244-12.2012
- Pytkäinen, L., Brennan, J., and Bemis, D. K. (2011). Grounding the cognitive neuroscience of semantics in linguistic theory. *Lang. Cogn. Process.* 26, 1317–1337. doi: 10.1080/01690965.2010.527490
- Saddy, D., Drenhaus, H., and Frisch, S. (2004). Processing polarity items: contrastive licensing costs. *Brain Lang.* 90, 495–502. doi: 10.1016/S0093-934X(03)00470-X
- Shao, J., and Neville, H. (1998). Analyzing semantic processing using Event-related Potentials. *News.Center Res. Lang.* 11, 3–20.
- Steinhauer, K., Drury, J. E., Portner, P., Walenski, M., and Ullman, M. T. (2010). Syntax, concepts and logic in the temporal dynamics of language comprehension. *Neuropsychologia* 48, 1525–1542. doi: 10.1016/j.neuropsychologia.2010.01.013
- Stowe, L. (1986). Parsing *wh*-constructions: evidence for on-line gap location. *Lang. Cogn. Process.* 1, 227–245. doi: 10.1080/01690968608407062
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *J. Mem. Lang.* 48, 542–562. doi: 10.1016/S0749-596X(02)00536-3
- Townsend, D. J., and Bever, T. G. (2001). *Sentence Comprehension: The Integration of Habits and Rules*. Cambridge, MA: MIT Press.
- Tsai, P.-S., Tzeng, O. J.-L., Hung, D. L., and Wu, D. H. (2013). Using magnetoencephalography to investigate processing of negative polarity items in Mandarin Chinese. *J. Neuroling.* 26, 258–270. doi: 10.1016/j.jneuroling.2012.09.002
- Tsai, W.-T. D. (1994). *On Economizing the Theory of A-bar dependencies*. Ph.D. dissertation, Cambridge; Boston, MA: MIT Press.
- Van Berkum, J. J. A., Koornneef, A. W., Ottena, M., and Nieuwland, M. S. (2007). Establishing reference in language comprehension: an electrophysiological perspective. *Brain Res.* 1146, 158–171. doi: 10.1016/j.brainres.2006.06.091
- van Gompel, R. P. G., and Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 128–139. doi: 10.1037/0278-7393.29.1.128
- Vasishth, S., Brüessow, S., Lewis, R. L., and Drenhaus, H. (2008). Processing polarity: how the ungrammatical intrudes on the grammatical. *Cogn. Sci.* 32, 685–712. doi: 10.1080/03640210802066865
- Vespignani, F., Panizza, D., Zandomenighi, P., and Job, R. (2009). ““Never” in the wrong place: an ERP study of unlicensed NPIs in Italian,” in *Poster Presented at the 22nd Annual CUNY Conference on Human Sentence Processing* (Davis).
- Warren, T., and Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition* 85, 79–112. doi: 10.1016/S0010-0277(02)00087-2
- Xiang, M., Grove, J., and Giannakidou, A. (2016). Semantic and pragmatic processes in the comprehension of negation: an event related potential study of negative polarity items. *J. Neuroling.* 38, 71–88. doi: 10.1016/j.jneuroling.2015.11.001
- Xiang, M., Dillon, B., and Phillips, C. (2009). Illusory licensing effects across dependency types. *Brain Lang.* 108, 40–55. doi: 10.1016/j.bandl.2008.10.002
- Xiang, M., Dillon, B., Wagers, M., Liu, F. Q., and Guo, T. M. (2013). Processing covert dependencies: an SAT study on Mandarin *wh-in-situ* questions. *J. East Asian Ling.* 23, 207–232. doi: 10.1007/s10831-013-9115-1
- Xiang, M., Wang, S. P., and Cui, Y. L. (2015). Constructing covert dependencies—the case of *wh-in-situ* processing. *J. Mem. Lang.* 84, 139–166. doi: 10.1016/j.jml.2015.05.006
- Xu, X., Jiang, X., and Zhou, X. (2013). Processing biological gender and number information during Chinese pronoun resolution: ERP evidence for functional differentiation. *Brain Cogn.* 81, 223–236. doi: 10.1016/j.bandc.2012.11.002
- Yanilmaz, A., and Drury, J. E. (2013). “Processing of negative polarity items in Turkish,” in *Poster Presented at the Society for the Neurobiology of Language* (San Diego, CA).
- Yen, H.-L. (2007). *Processing of Proper Names in Mandarin Chinese: A Behavioral and Neuroimaging Study*. Ph.D. dissertation, Uni Bielefeld.
- Yoshida, M., Kazanina, N., Pablos, L., and Sturt, P. (2014). On the origin of islands. *Lang. Cogn. Process.* 29, 761–770. doi: 10.1080/01690965.2013.788196
- Yurchenko, A., den Ouden, D. B., Hoeksema, J., Dragoy, O., Hoeks, J. C. J., and Stowe, L. A. (2013). Processing polarity: ERP evidence for differences between positive and negative polarity. *Neuropsychologia* 51, 132–141. doi: 10.1016/j.neuropsychologia.2012.10.028

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer, EC, and handling Editor declared their shared affiliation.

Copyright © 2018 Pablos, Doetjes and Cheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# The Limited Role of Number of Nested Syntactic Dependencies in Accounting for Processing Cost: Evidence from German Simplex and Complex Verbal Clusters

Markus Bader\*

Department of Linguistics, Goethe University Frankfurt, Frankfurt am Main, Germany

## OPEN ACCESS

### Edited by:

Ángel J. Gallego,  
Universitat Autònoma de Barcelona  
Spain

### Reviewed by:

Cristiano Chesi,  
Istituto Universitario di Studi Superiori  
di Pavia (IUSS), Italy  
Leticia Pablos,  
Leiden University Centre for  
Linguistics (LUCL), Netherlands

### \*Correspondence:

Markus Bader  
bader@em.uni-frankfurt.de

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 30 April 2017

**Accepted:** 13 December 2017

**Published:** 23 January 2018

### Citation:

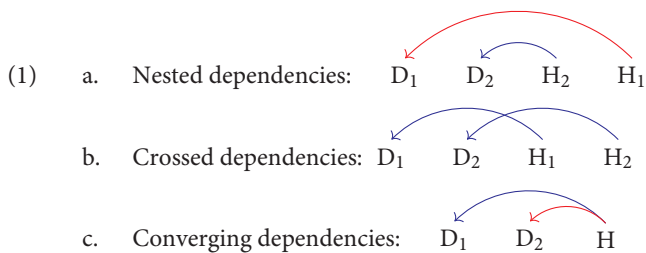
Bader M (2018) The Limited Role of  
Number of Nested Syntactic  
Dependencies in Accounting for  
Processing Cost: Evidence from  
German Simplex and Complex Verbal  
Clusters. *Front. Psychol.* 8:2268.  
doi: 10.3389/fpsyg.2017.02268

This paper presents three acceptability experiments investigating German verb-final clauses in order to explore possible sources of sentence complexity during human parsing. The point of departure was De Vries et al.'s (2011) generalization that sentences with three or more crossed or nested dependencies are too complex for being processed by the human parsing mechanism without difficulties. This generalization is partially based on findings from Bach et al. (1986) concerning the acceptability of complex verb clusters in German and Dutch. The first experiment tests this generalization by comparing two sentence types: (i) sentences with three nested dependencies within a single clause that contains three verbs in a complex verb cluster; (ii) sentences with four nested dependencies distributed across two embedded clauses, one center-embedded within the other, each containing a two-verb cluster. The results show that sentences with four nested dependencies are judged as acceptable as control sentences with only two nested dependencies, whereas sentences with three nested dependencies are judged as only marginally acceptable. This argues against De Vries et al.'s (2011) claim that the human parser can process no more than two nested dependencies. The results are used to refine the Verb-Cluster Complexity Hypothesis of Bader and Schmid (2009a). The second and the third experiment investigate sentences with four nested dependencies in more detail in order to explore alternative sources of sentence complexity: the number of predicted heads to be held in working memory (storage cost in terms of the Dependency Locality Theory [DLT], Gibson, 2000) and the length of the involved dependencies (integration cost in terms of the DLT). Experiment 2 investigates sentences for which storage cost and integration cost make conflicting predictions. The results show that storage cost outweighs integration cost. Experiment 3 shows that increasing integration cost in sentences with two degrees of center embedding leads to decreased acceptability. Taken together, the results argue in favor of a multifactorial account of the limitations on center embedding in natural languages.

**Keywords:** syntactic dependencies, processing complexity, center embedding, recursion, verb cluster, German

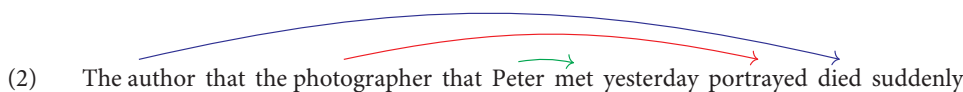
# 1. INTRODUCTION

One of the few features of natural languages for which there is general agreement is the existence of non-local dependencies (Tallerman et al., 2009). Within psycholinguistics, non-local dependencies play a key role in several theories of the human parser, including the Syntactic Prediction Locality Theory (Gibson, 1998), the Dependency Locality Theory (Gibson, 2000), the Efficiency Theory (Hawkins, 2004, 2014), and the Minimize Dependencies Theory (Temperley, 2007; Gildea and Temperley, 2010). These theories have focused on two properties of syntactic dependencies, their number and their length, but syntactic dependencies have other properties which may be relevant too. The way dependencies are ordered is such a property, as pointed out by De Vries et al. (2011). Two overlapping dependencies can be ordered in one of three ways, as shown in (1), where  $D_n$  is dependent on  $H_{(n)}$ .



First, one dependency can be nested within the other, as in (1-a); second, two dependencies can cross each other, as in (1-b); third, two dependencies can converge on a single head, as in (1-c).

It has often been observed that sentences with multiple center embedding are difficult or even impossible to comprehend (see reviews in Gibson, 1998; De Vries et al., 2011). Even sentences with only two levels of center embedding and thus three nested dependencies, as illustrated in (2), can be difficult for the human parser to process. The reasons for this limitation on human parsing are still a matter of active research.



Sentences with doubly center embedded relative clauses are the most prominent instance of multiply nested dependencies, but they are not the only ones. As pointed out by De Vries et al. (2011), another instance is provided by sentences with certain types of complex verb clusters as they are found in the West-Germanic verb-final languages, including Dutch and German. Experimental evidence on this issue comes from a seminal study on crossed and nested dependencies by Bach et al. (1986). Bach et al. (1986) capitalized on the fact that Dutch and German, despite being syntactically highly similar, differ with regard to the order of verbs. When several verbs appear in a row in clause-final position, they form a so-called *verb cluster*. As illustrated in **Table 1**, verb clusters give rise to crossed dependencies in Dutch but nested dependencies in German.

**Table 1** shows only a subset of all dependencies in the sentences under consideration. The dependencies that are shown

**TABLE 1 |** Order of dependencies in Dutch and German sentences with 2 and 3 verb clusters.

2-verb cluster	German	dass Jan Maria schwimmen sah that Jan.NOM Maria.ACC swim saw
	Dutch	dat Jan Maria zag zwemmen that Jan.NOM Maria.ACC saw swim
3-verb cluster	German	dass Jan Peter Maria schwimmen lassen sah that Jan.NOM Peter.ACC Maria.ACC swim let saw
	Dutch	dat Jan Piet Maria zag laten zwemmen that Jan.NOM Peter.ACC Maria.ACC saw let swim

are those between verbs and their NP arguments, that is, those dependencies which are necessary for assigning semantic roles. For example, the NP *Jan* is the subject argument of *sag/sah* (“saw”). The dependencies between a verb and its verbal arguments, for example between *sag/sah* (“saw”) and *zwemmen/schwimmen* (“swim”) in the 2-verb cluster sentences, are not shown because these are all local dependencies which do not contribute to the issue of how nested dependencies affect sentence complexity.

Bach et al. (1986) had speakers of Dutch and German rate sentences as shown in **Table 1** in their respective language. Verb clusters of size one to four were included in the study. Bach et al.’s (1986) experiment yielded two major results. First, sentences with two nested or crossed dependencies showed only a small decrease in acceptability in comparison to sentences with only a single dependency, but adding a third dependency caused acceptability to decline sharply. Going from two verbs to three verbs decreased acceptability by about three points on a scale from 1 to 10, and going from three verbs to four verbs led to a further decrease of 2 points. Order of dependencies did not have a significant effect for two dependencies, but for three or four dependencies, acceptability declined less sharply for crossed than for nested dependencies. On average, an advantage of about 0.4 points was observed when comparing crossed dependencies

to nested dependencies for clusters of equal size. In sum, clusters of size three or greater are hardly processable whether they involve crossed or nested dependencies, and the disadvantage is somewhat stronger for nested than for crossed dependencies.

Based on Bach et al.’s (1986) finding as well as on evidence concerning multiply center-embedded relative clauses, De Vries et al. (2011) arrive at the conclusion that ...

[...] since humans possess finite brains that are constrained by (among other things) memory limitations, we have problems comprehending and producing sentences with three or more nested or crossed dependencies [...]  
 (De Vries et al., 2011, p. 12)

De Vries et al. (2011) put forward an interesting generalization which delimits the class of sentences leading to processing overload in an empirically testable way. In order to test this

hypothesis and to explore the role that dependency formation may play for sentence complexity more generally, this paper presents three experiments that have investigated German verb-final clauses of varying complexity. The first experiment provides a test of De Vries et al.'s (2011) generalization. Since the results of this test show that the generalization is not correct, two further experiments explore alternative sources of sentence complexity, namely integration and storage cost as defined in the Dependency Locality Theory of Gibson (2000). Before the experiments are presented, the next section gives a short introduction to current accounts of parsing complexity.

## 2. DETERMINANTS OF SYNTACTIC COMPLEXITY

Memory and expectations are the main ingredients of current theories of syntactic processing complexity (see Nakatani and Gibson, 2008; Jaeger and Tily, 2011; Levy, 2013). With regard to memory, syntactic dependency formation during on-line sentence comprehension poses several requirements. When the two elements of a dependency relation occur adjacent to each other, the first element is still in the focus of attention and thus immediately available for being integrated with the second element. In the case of non-local dependencies, however, the first and second element of a dependency are separated by intervening material. In this case, the element of the dependency that comes first in the word string must be kept in working memory for later retrieval on encountering the second element. Keeping elements in memory and retrieving elements from memory are both possible sources of sentence complexity. The idea that parsing complexity varies with the number of dependencies for which the first element has already been encountered but not the second element, has been termed the *Incomplete Dependency Hypothesis* in Nakatani and Gibson (2008). This hypothesis is given in (3).

### (3) *Incomplete Dependency Hypothesis*

The human sentence processor is sensitive to the number of partially processed dependencies at each processing state.

For the case of converging dependencies, that is, dependencies in which several phrases are dependent on a single head, research on verb-final languages has repeatedly shown that increasing the number of incomplete dependencies does not lead to increased processing load and can make a sentence even less difficult to process. For example, Nakatani and Gibson (2008) ran a self-paced reading experiment investigating Japanese sentences with one degree of center embedding and a varying number of incomplete dependencies. Two conditions from this experiment are illustrated in (4).

- (4) denwaban-ga                      sin'nyuusyain-ga (kokyaku-ni) tyuumonsyo-o hassoosita to dentatusita ato  
 telephone receptionist.NOM freshman.NOM client.DAT order sheet.ACC sent that told after  
 'After the telephone receptionist told (somebody) that the freshman had sent the order sheet to the client, ...'

In (4), the higher temporal clause contains a subject and a complement clause. This complement clause in turn contains a subject, an accusative object, and optionally a dative object. Directly before encountering the verb of the embedded complement clause, there are three incomplete dependencies when the dative object is absent (the subject of the matrix clause and the subject and accusative object of the embedded clause) and four when the dative object is present. Despite the high number of incomplete dependencies, such sentences do not pose problems for the human parser and reading times were in fact lower in the presence of a dative object, that is, with four instead of three incomplete dependencies. Thus, instead of making sentences difficult to comprehend, a high number of incomplete dependencies can ease sentence processing. This effect, which was first found by Konieczny (2000), has become known as the *anti-locality effect* (see Vasishth and Lewis, 2006; Levy and Keller, 2013, for related findings).

In sum, there is abundant evidence showing that increasing the number of incomplete converging dependencies does not increase processing load. For nested and crossed dependencies, the situation is less clear. A close relationship between the number of nested dependencies and parsing complexity is suggested by sentences with multiple center embedding. As illustrated by example (2), sentences with two levels of center embedding contain three nested dependencies, and such sentences are difficult to process. Similar considerations hold for verb clusters with three or more verbs as investigated by Bach et al. (1986). Findings of this kind have led De Vries et al. (2011) to the generalization that processing three or more nested or crossed dependencies is beyond the normal capacity of the human parser.

As already pointed out above, non-local dependencies require not only to keep the first element in memory until the second element is encountered. They also require to retrieve the first element from the ongoing memory representation on encountering the second element. Retrieval may be difficult because the two elements are separated by intervening material. Of particular importance in this regard is the distance between the two elements of a dependency, that is, the length of the dependency. This aspect of dependency formation has been termed the *Bottom-up Head-dependent Distance Hypothesis* by Nakatani and Gibson (2008). This hypothesis is given in (5).

### (5) *Bottom-up Head-dependent Distance Hypothesis*

The difficulty of integrating a new word *w* into the current structure depends on the distance back to the head *h* to which *w* connects.

The distance between the two elements of a dependency can be measured in different ways. For the following discussion, dependency length is measured in terms of integration cost as proposed in the Dependency Locality Theory of Gibson (2000). The (total) integration cost of a word is the sum of its referential integration cost and its structural integration cost. Each word triggering the introduction of a new discourse referent is assumed to incur a referential integration cost of one unit, where nouns and verbs introduce new discourse referents and all other words do not (see Warren and Gibson, 2002, for a more fine-grained measure). Structural integration cost arises when a dependency has to be formed, that is, when a new input word must be integrated with a word already contained within sentence memory. Structural integration cost is a function of dependency length, with length measured in terms of the number of new discourse referents that intervene between the two items of a dependency.

Assigning a syntactic structure to a sentence during parsing involves more than computing the various dependencies that obtain between the words and phrases of the sentence. In particular, the human parser also has to compute a phrase-structure representation. This task provides a possible further source of sentence complexity. As long as the phrase-structure representation is not complete, the parser may form expectations about how the partial phrase-structure tree computed for the input string seen so far will be completed by the remainder of the input string. As in the case of incomplete dependencies, the simplest way to link phrase-structure expectations to sentence complexity is by counting the number of expectations that have to be held in working memory at each point during the ongoing parse. One implementation of this idea, which goes back to the early work of Yngve (1960) on language production, is stated in the *Predicted Syntactic Head Hypothesis* of Nakatani and Gibson (2008) given in (6).

(6) *Predicted Syntactic Head Hypothesis*

The human sentence processor is sensitive to the number of syntactic heads that are required to form a grammatical sentence at each processing state.

In accordance with the DLT's notion of storage cost, it is assumed below that each predicted head is associated with one memory unit. In the following, only the storage cost associated with predicted verbal heads is considered, because storage cost related to other heads is always matched across sentences compared to each other.

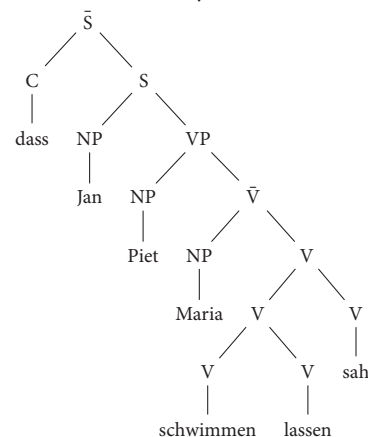
To summarize, number of nested or crossed dependencies, the length of the dependencies, and the ongoing phrase-structure representation have been proposed as potential sources of sentence complexity. These sources do not exclude each other and they are not meant as an exhaustive list. The following three experiments were designed to test the contribution of each of the three sources of complexity. Whether the number of nested dependencies affects sentence

complexity is the topic of Experiment 1. Experiment 2 tests the relative importance of integration cost and storage cost in sentences where the two make divergent predictions. The final Experiment 3 investigates the role of integration cost in sentences matched for number of nested dependencies and storage cost.

### 3. EXPERIMENT 1

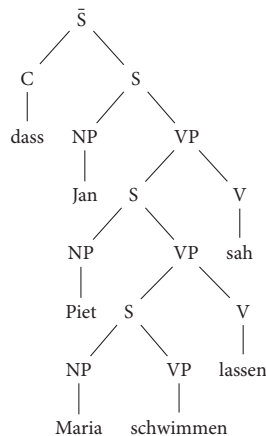
The aim of Experiment 1 was to test De Vries et al.'s (2011) hypothesis that sentences with three or more crossed or nested dependencies cause problems for the human parsing mechanism. Experiment 1 tests this hypothesis by disentangling number of nested dependencies and verb cluster complexity. Verb cluster formation is a typologically rare property of the Germanic OV languages (see Wurmbrand, 2006, 2017, for comprehensive overviews). As documented by Wurmbrand (2006), syntactic analyses of verb clusters differ in many important ways from each other (see Seuren and Kempen, 2003, for a selection of verb cluster analysis within a broad range of syntactic frameworks). In the current context, the most important property of verb clusters is that in many respects, they behave like a single verbal head, independent of the number of verbs they contain. Thus, a sentence with a 3-verb cluster would get a surface structure along the lines of the mono-clausal representation in (7)<sup>1</sup>. Such a mono-clausal structure may be syntactically derived from a multi-clausal structure as in (8), but it can also be generated directly and without reference to any kind of multi-clausal syntactic structure. A multi-clausal representation may still be necessary, but only at the semantic level.

(7) Mono-clausal analysis of verb clusters



<sup>1</sup>For reasons of simplicity, the phrase-structure trees in (7) and (8) are expressed in terms of  $\bar{S}$ , S and VP. Using a more articulated structure involving, for example, CP, IP, and VP, wouldn't change the argument.

## (8) Multi-clausal analysis of verb clusters



Verb clusters with three or more verbs in a row are not necessarily hard to process for the human parser. When only a single verb introduces arguments into the clause, verb clusters up to five verbs can be comprehended without much difficulty, and verb clusters of this size occur in authentic texts, as shown in (9) (see Bader and Schmid, 2009b; Bader et al., 2009, for experimental evidence and corpus data).

- (9) ... was alles besser **hätte**<sub>1</sub> gemacht<sub>5</sub> worden<sub>4</sub> sein<sub>3</sub>  
 what all better had mad been be  
 können<sub>2</sub>  
 can  
 ‘what could have been made better’  
 (www.dradio.de/dkultur/sendungen/fazit/2028303/)

With regard to the relationship between verb-cluster formation and sentence complexity, the empirical data can be summarized as follows. The data of Bach et al. (1986) indicate that verb clusters in which each verb introduces its own argument(s) are easy to process as long as no more than two verbs are involved. With three or more verbs, such clusters become difficult or even impossible to comprehend. Adding further verbs not introducing arguments of their own, in contrast, increases complexity only marginally if at all.

A possible source of the processing complexity observed in the case of Bach et al.’s (1986) sentences is verb-cluster formation itself. A proposal to this effect has been made by Bader and Schmid (2009a), based on an investigation of so-called *long passivization*, as illustrated by the example in (10).

- (10) Es wurde berichtet, dass der alte Vater  
 [zu entlasten versucht] wurde.  
 it was reported that the.NOM old father  
 to disburden tried was  
 ‘It was reported that one had tried to disburden the old father.’

Here the control verb *versuchen* (“to try”) occurs in the passive voice, as shown by its appearance as a past participle.

The unexpected property of this construction is that the major change brought about by passivization, the promotion of the direct object to subject, does not affect the object of the passivized verb *versuchen* (“try”), but the object of the infinitival verb *zu entlasten* (“to disburden”), which is the complement of *versuchen* (“to try”). This object occurs with nominative case in (10) instead of accusative case, as it would in a corresponding active clause. Passivization thus has a kind of long-distance effect in this construction, hence the name long-distance passivization. However, if *zu entlasten* (“to disburden”) and *versuchen* (“to try”) form a verb cluster and thus a single complex predicate, passivization applies in the usual way. What is passivized is not *versuchen* (“to try”) itself, but *zu entlasten versuchen* (“to try to disburden”) as a whole. As shown by the somewhat reduced acceptability of this construction, forming a complex predicate and then applying passivization to it is not cost-free. Bader and Schmid (2009a) have therefore proposed the *Verb-Cluster Complexity Hypothesis* given in (11).

- (11) *Verb-Cluster Complexity Hypothesis*  
 The argument-structure operations involved in verb-cluster formation are costly for the HSPM [= Human Sentence Processing Mechanism].

The Verb-Cluster Complexity Hypothesis was stated under the assumption that only verbs that have arguments of their own come with an argument structure. These are all lexical verbs, whereas functional verbs like auxiliaries and modals have no arguments. A small number of verbs have a hybrid status, like the verb *lassen* (“to let”), which has a causer argument but shows the syntactic behavior of modal verbs. Under these assumptions, the Verb-Cluster Complexity Hypothesis distinguishes between verb clusters that involve only a single verb with an argument structure and verb clusters in which the argument structures of several verbs must be combined in some way. What this hypothesis does not predict is why there is a rather sharp decline in acceptability when more than two argument structures must be combined, as in the 3- and 4-verb clusters investigated by Bach et al. (1986).

A major drawback of the *Verb-Cluster Complexity Hypothesis* is that it is specifically tailored to the case of verb-cluster formation. This contrasts with De Vries et al.’s (2011) account, which derives the complexity observed for clusters with three or more verbs from a general constraint on human parsing, namely that parsing proceeds smoothly only when a sentence contains no more than two crossed or nested dependencies. This generalization predicts that three or more nested dependencies should cause high processing complexity independently of whether a complex verb cluster is involved or not. This prediction can be tested with the help of sentences in which three or more nested dependencies are distributed across several verb clusters with at most two verbs. Two examples with three nested dependencies distributed across two verb clusters—a 1-verb cluster and a 2-verb cluster—are shown in (12).

- (12) a. dass der Lehrer, der Maria schwimmen ließ, lachte  
 that the teacher.NOM who.NOM Maria.ACC swim let laughed  
 'that the teacher who let Maria swim laughed.'
- b. dass Jan den Lehrer, der schwamm, lachen sah  
 that Jan.NOM the.ACC teacher who.NOM swam laugh saw  
 'that Jan saw the teacher who swam laugh.'

When both the upper and the lower clause contain a 2-verb cluster, four nested dependencies result, as shown in (13).

- (13) dass Jan den Lehrer, der Maria schwimmen ließ, lachen sah  
 that Jan.NOM the.ACC teacher who.NOM Maria.ACC swim let laugh saw  
 'that Jan saw the teacher who let Maria swim laugh.'

If it were true that sentences with three or more nested dependencies exceed the normal capacity of the human parsing mechanism, sentences as in (12) and (13) should be at least as difficult to process than the three- and four-verb cluster sentences investigated by Bach et al. (1986). If, on the other hand, the findings of Bach et al. (1986) reflect processing complexity tied to verb-cluster formation itself, then sentences containing three or more nested dependencies should become easier to process when they do not contain a complex verb cluster. These predictions are tested in Experiment 1 by comparing the complexity of sentences containing three nested dependences originating in a single 3-verb cluster (1×3 sentences) to the complexity of sentences with four nested dependencies distributed across two verb clusters with two verbs each (2×2 sentences). Complexity will be assessed using an acceptability rating task instead of an on-line measure in order to obtain results that are comparable to the results of Bach et al. (1986).

1×3 sentences are structurally similar to the 3-verb cluster sentences investigated by Bach et al. (1986). Given their results, 1×3 sentences are expected to be of marginal acceptability. If De Vries et al. (2011) are correct and it is the presence of three nested dependencies which makes 1×3 sentences difficult to comprehend, then 2×2 sentences, which contain four nested dependencies, should be even less acceptable. If, on the other hand, the complexity of 1×3 sentences is intimately tied to the presence of a 3-verb cluster, 2×2 sentences should be more acceptable than 1×3 sentences.

### 3.1. Methods

#### 3.1.1. Participants

Sixty-four students from the Goethe-University Frankfurt completed a questionnaire for course credit. All participants were native speakers of German and naive with respect to the purpose of the experiment. Ethical approval was not required for this study in accordance with the national and institutional guidelines.

#### 3.1.2. Materials

Sixteen sentences were constructed for Experiment 1. Each sentence appeared in four versions according to the two factors

Dependencies (1×3 vs. 2×2) and Structure (center embedded vs. control). Center embedded 1×3 sentences were included to replicate the finding that three-verb clusters as investigated by Bach et al. (1986) are difficult to comprehend. Bach et al. are not very explicit concerning their experimental material and give only an example sentence representing their three-verb cluster condition. This sentence is reproduced in (14).

- (14) Arnim hat Wolfgang der Lehrerin  
 Arnim.NOM has Wolfgang.ACC the.DAT teacher  
 die Murmeln aufräumen helfen lassen.  
 the.ACC marbels collect-up help let  
 Arnim let Wolfgang help the teacher collect up the  
 marbles.

In addition to containing a complex verb cluster, sentence (14) is complex in several other ways. First, because this sentence is a main clause, a composite tense form with the finite auxiliary in the verb-second position must be used in order to have three verbs in clause-final position. Since for some of the verbs used by Bach et al. (1986) there was an uncertainty with regard to the morphological form required in the perfect tense (past participle or infinitive), the authors ran two subexperiments varying the morphological form. Second, this sentence is more complicated than the sentences considered so far because it contains four arguments instead of three, which is a consequence of using the control verb *helfen* ("to help"), which has a dative object in addition to its verbal complement. Third, *helfen* ("to help") can also be used with a *zu* ("to") infinitive instead of a bare infinitive, introducing some indeterminacy that is not found with other verbs selecting a verbal complement.

Since these complications may have contributed to the reduced acceptability of sentences with complex verb clusters in Bach et al. (1986), Experiment 1 investigated center embedded 1×3 sentences that differed in several ways from sentences as in (14). First of all, the complex verb cluster was contained within an embedded verb-final clause in Experiment 1. As a

consequence, the finite verb appeared clause-finally and it was no longer necessary to use a composite tense form. Instead, all verb clusters ended with a main verb in the past tense. Second, only three verbs selecting a verbal complement were used. The hierarchically highest and thus the finite verb was always the verb *sah* (“saw”), which is the most frequent perception verb. This verb selected either the verb *lassen* (“let”) or the verb *versuchen* (“try”), which is the most acceptable control verb in this kind of construction (Schmid et al., 2005). All three verbs unambiguously determine the morphological form of their verbal complement.

Using the three verbs *sah* (“saw”), *lassen* (“let”) and *versuchen* (“try”), three types of sentences were constructed. All three sentence types instantiate the structural pattern described in the introduction to Experiment 1, but differ in several item-specific ways. If the complexity of these sentences in the four different experimental conditions is mainly driven by structural factors, all three sentence types are expected to show the same pattern of acceptability. Should Experiment 1 reveal distinct acceptability patterns for the three sentence types, the assumption that complexity is a function of sentence structure would have to be abandoned.

An example sentence for each of the three sentence types is shown in **Table 2**; the complete sentence set is available as Supplementary Material. All sentences consisted of the main clause “Ich weiß” (“I know”) followed by a *that*-clause. In the condition “center-embedded with 1×3 dependencies,” the *that*-clause contained three NPs followed by a 3-verb cluster. The first NP was a proper name and the second and third NP were definite NPs marked for accusative case. Eight sentences contained a 3-verb cluster of the form “lexical verb – *lassen* (“let”) – *sah* (“saw”)”. In four of the sentences with *lassen*, the lexical verb was intransitive and the third NP realized the subject argument of this verb, as in the example discussed above. In the other four sentences with *lassen*, the lexical verb was a transitive verb and its object was realized by the third NP. In this case, the subject of the lexical verb is implicitly understood as “someone.”

The remaining eight sentences contained a verb cluster of the form “lexical verb – *versuchen* (“to try”) – *sah* (“saw”)”. In these sentences, the third NP was the object of the lexical verb, which always was a transitive verb, and the subject of the lexical verb was implicitly understood as the subject of the control verb *versuchen* (“to try”).

Sentences in the condition “center-embedded with 2×2 dependencies” were created from sentences in the condition “center-embedded with 1×3 dependencies” as follows. First, the *that*-clause now contained only the first two NPs, the subject and the first accusative NP. This NP was modified by a relative clause introduced by a subject relative pronoun. This relative clause contained the former second accusative NP. The relative clause ended in a 2-verb cluster containing the lexical verb and the second verb of the original 3-verb cluster. All *that*-clauses also ended in a 2-verb cluster with the verb *sah* (“saw”) as finite verb. The non-finite verb of the verb cluster was a lexical verb that did not occur in corresponding 1×3 sentences.

In this and the following two experiments, control sentences were derived from the experimental sentences by means of extraposition, thereby eliminating center embedding or at least

reducing it. For the 1×3 sentences, control sentences were derived by extraposing the complement of the finite perception verb *sah* (“saw”), that is, the two infinitive verbs embedded below *saw* together with their arguments. Because a perception verb can take an infinitival complement only when this complement occurs to its left, the extraposed clause had to be turned into a finite clause introduced by *wie* (“how”). Despite this morpho-syntactic difference, the control sentences had the same meaning as the experimental sentences with center embedding. In the condition “2×2 dependencies,” the relative clause was extraposed behind the verb cluster of the *that*-clause. In this case, the extraposed clause had not to be modified in any way. Experimental and control sentences were thus identical with the exception of the position of the relative clause.

The 16 sentence quadruples were distributed onto four experimental lists according to a Latin square design. Each experimental list contained only one version of each sentence, with an equal number of sentences occurring in each of the four experimental conditions. Each experimental list was randomized and then combined with a list of 72 filler sentences. The filler sentences represented a variety of sentence structures and were partly taken from unrelated experiments.

### 3.1.3. Procedure

Four written questionnaires were produced on the basis of the four lists of experimental and filler sentences. Participants completed the questionnaires as part of a class session. They were asked to judge the acceptability of each item on the questionnaire by marking one of the numbers 1 to 7 printed beneath each sentence. A scale ranging from 1 to 7 was chosen because such a scale is in common use (Schütze and Sprouse, 2014) and has proved its usefulness in numerous experiments (e.g., Weskott and Fanselow, 2011). A short instruction on the first page of the questionnaire told participants that 1 meant “totally unacceptable” and 7 meant “totally acceptable” (see the Supplementary Material for the complete instruction). The instruction did not contain any example sentences. Participants needed about 15–20 min to complete the questionnaire.

## 3.2. Results

All data presented in this paper were analyzed using the R statistics software, Version 3.3.2 (R Core Team, 2016). To test for significant effects, the judgment data were analyzed by means of mixed-effect modeling using the lme4 package (Bates et al., 2015b). The experimental factors and all interactions between them were entered as fixed effects into the model, using effect coding, that is, the intercept represents the unweighted grand mean and fixed effects compare factor levels to each other. In addition, the model included random effects for items and subjects with maximal random slopes supported by the data, following the strategy proposed in Bates et al. (2015a). The full model summary is reported as well as likelihood ratio tests, which assess the contribution of single factors or interactions. Where necessary, pairwise comparisons were computed using Tukey’s test.

**Figure 1** shows the mean acceptability ratings for the three sentence types investigated in Experiment 1. The basic

**TABLE 2** | Three complete stimulus sentences from Experiment 1, one with a causative verb and one with a control verb.

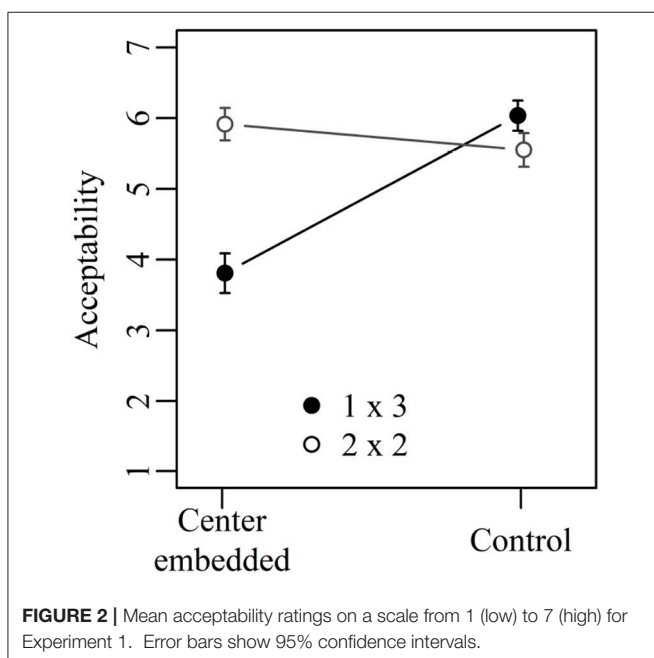
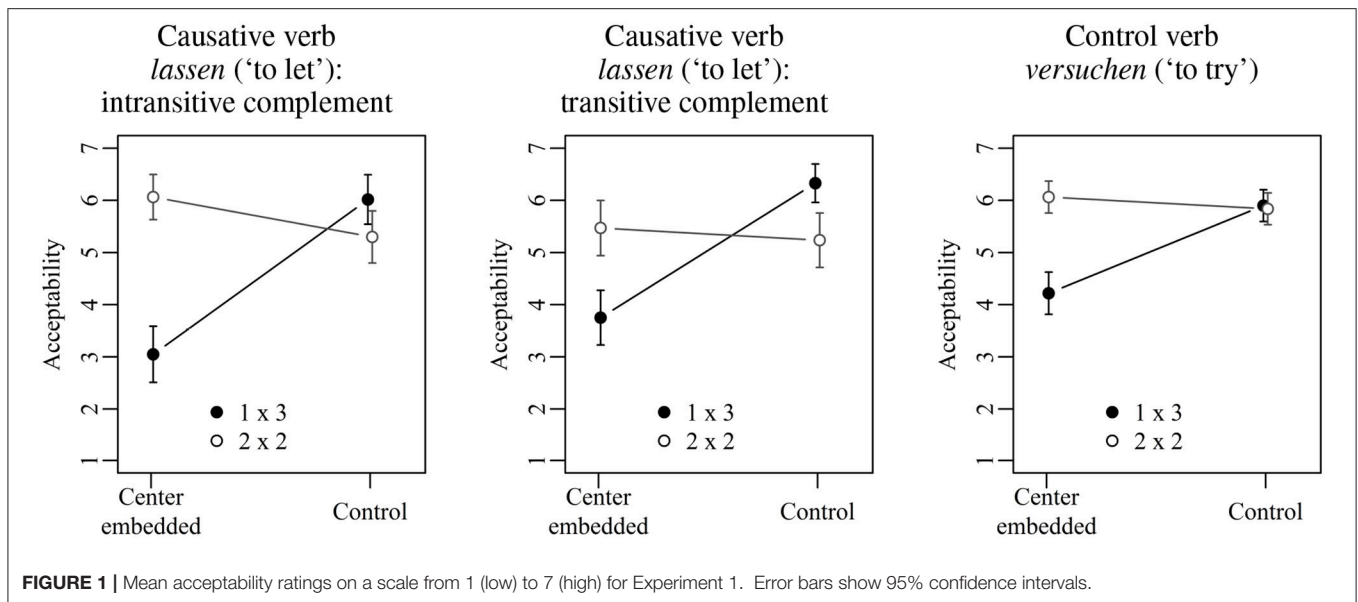
CAUSATIVE VERB LASSEN TO LET – INTRANSITIVE COMPLEMENT		
1 × 3	center	dass Moritz den Gärtner den Kunden warten lassen sah. that M.NOM the.ACC gardener the.ACC costumer wait let see 'that Moritz saw the gardener letting the costumer wait.'
	embedded	
	control	dass Moritz sah, wie der Gärtner den Kunden warten ließ. that M.NOM saw how the.NOM gardener the.ACC costumer wait let 'that Moritz saw how the gardener let the costumer wait.'
2 × 2	center	dass Moritz den Gärtner, der den Kunden warten ließ, arbeiten sah. that M.NOM the.ACC gardener who.NOM the.ACC costumer wait let work saw 'that Moritz saw the gardener that let the costumer wait work.'
	embedded	
	control	dass Moritz den Gärtner arbeiten sah, der den Kunden warten ließ. that M.NOM the.ACC gardener work saw who.NOM the.ACC costumer wait let 'that Moritz saw the gardener work that let the costumer wait.'
CAUSATIVE VERB LASSEN TO LET – TRANSITIVE COMPLEMENT		
1 × 3	center	dass Alexander den König den Dieb bestrafen lassen sah. that A.NOM the.ACC king the.ACC thief punish let saw 'that Alexander saw the king letting punish the thief.'
	embedded	
	control	dass Alexander sah, wie der König den Dieb bestrafen ließ. that A.NOM saw how the.NOM king the.ACC thief punish let 'that Alexander saw how the king let punish the thief.'
2 × 2	center	dass Alexander den König, der den Dieb bestrafen ließ, lachen sah. that A.NOM the.ACC king who.NOM the.ACC thief punish let laugh saw 'that Alexander saw the king that let punish the thief laugh.'
	embedded	
	control	dass Alexander den König lachen sah, der den Dieb bestrafen ließ. that A.NOM the.ACC king laugh saw who.NOM the.ACC thief punish let 'that Alexander saw the king laugh that let punish the thief.'
CONTROL VERB VERSUCHEN TO TRY		
1 × 3	center	dass Peter den Koch den Brand zu löschen versuchen sah. that P.NOM the.ACC cook the.ACC fire to extinguish trying saw 'that Peter saw the cook trying to extinguish the fire.'
	embedded	
	control	dass Peter sah, wie der Koch den Brand zu löschen versuchte. that P.NOM saw how the.NOM cook the.ACC fire to extinguish tried 'that Peter saw how the cook tried to extinguish the fire.'
2 × 2	center	dass Peter den Koch, der den Brand zu löschen versuchte, verzweifeln sah. that P.NOM the.ACC cook who.NOM the.ACC fire to extinguish tried despair saw 'that Peter saw the cook that tried to extinguish the fire despair.'
	embedded	
	control	dass Peter den Koch verzweifeln sah, der den Brand zu löschen versuchte. that P.NOM the.ACC cook despair saw who.NOM the.ACC fire to extinguish tried 'that Peter saw the cook despair that tried to extinguish the fire.'

All sentences were introduced by the main clause "Ich weiß" ("I know").

pattern is the same in each case: 1×3 sentences with center embedding receive much lower mean ratings than 1×3 control sentences. In the 2×2 condition, in contrast, sentences with center embedding are judged as equally or even slightly more acceptable than control sentences. Although the exact mean ratings differ somewhat across the three sentence types, an initial statistical analysis including sentence

type as a third factor showed neither a significant main effect of sentence type nor a significant interaction involving sentence type. The results thus do not depend on the specific combination of verbs with their associated lexical requirements, but on the more general structural configurations. The factor sentence type was accordingly dropped from the analysis.





**Figure 2** shows the mean acceptability ratings obtained in Experiment 1 collapsed across the three conditions of sentence type. The results of the corresponding statistical analysis are shown in **Table 3**. The two main effects as well as the interaction between them were significant. 1×3 sentences with center embedding received significantly lower acceptability ratings than 1×3 control sentences (3.8 vs. 6.0; Tukey's test:  $t$ -ratio = 10.21;  $p < 0.001$ ). The acceptability of 2×2 sentences with center embedding, in contrast, did not differ significantly from 2×2 control sentences (5.9 vs. 5.6; Tukey's test:  $t$ -ratio = 1.67,  $p > 0.1$ ). Furthermore, there was no significant acceptability

**TABLE 3 |** Linear mixed model fitted by maximum likelihood estimation for Experiment 1, including  $p$ -values from likelihood ratio tests.

	Coefficient	Std. Error	$t$ -value	$p$ (LRT)
(Intercept)	5.3271	0.1696	31.412	
Structure	-0.9316	0.1413	-6.592	<0.001
Dependencies	-0.8105	0.1684	-4.814	<0.001
Structure:dependencies	-2.5898	0.3322	-7.795	<0.001

$Acceptability \sim structure * dependencies + (dependencies || subject) + (structure * dependencies || sentence)$

difference when comparing 1×3 control sentences and 2×2 control sentences (6.0 vs. 5.6; Tukey's test:  $t$ -ratio = 2.05,  $p > 0.1$ ), whereas for center-embedded sentences, the corresponding comparison was significant (3.8 vs. 5.9; Tukey's test:  $t$ -ratio = 8.90;  $p < 0.001$ ).

### 3.3. Discussion

Experiment 1 yielded two major results. First, sentences with three nested dependencies all originating in a single 3-verb cluster are difficult to process. This replicates the original finding of Bach et al. (1986). The new finding of Experiment 1 is that 2×2 sentences are more acceptable than 1×3 sentences, and in fact no less acceptable than control sentences containing the same number of dependencies but with a maximum number of 2 nestings. This contradicts De Vries et al.'s (2011) generalization that sentences containing three or more nested dependencies pose special challenges to the human parser. Thus, the Incomplete Dependency Hypothesis in (3) is incorrect even if restricted to nested dependencies, which are all distinct in the sense of connecting each argument to a separate head.

Given the results of Experiment 1, the difficulty of the 3-verb clusters considered here cannot be attributed to general

limitations of the human parsing mechanism with regard to the processing of nested dependencies. This leaves us with the question of why verb clusters with more than two verbs lead to heavy processing load in cases where each verb introduces arguments of its own. The Verb-Cluster Complexity Hypothesis in (11) provides an answer to this question, but it is specifically tailored to the case of verb-cluster formation. It should therefore be accepted only if the findings of Experiment 1 cannot be accounted for by general theories of syntactic complexity.

In section 2, two general sources of sentence complexity were discussed in addition to the number of open nested dependencies, namely integration cost capturing dependency distance and storage cost capturing phrase-structure complexity. As shown in detail in the Supplementary Material, integration and storage cost do not provide an account for the low acceptability of Dutch and German verb clusters with 3 or more argument-taking verbs. This does not argue against the explanatory potential of these notions, but instead points to the conclusion that verb-cluster formation by itself can result in enhanced processing complexity under certain circumstances. In order to understand what makes verb clusters hard to process, the empirical findings concerning the processing complexity of sentences with verb clusters are summarized below.

- Verb clusters with only one argument-taking verb are (relatively) easy even if containing up to 5 verbs (Bader and Schmid, 2009b; Bader et al., 2009).
- Verb clusters with two argument-taking verbs are easy (Bach et al., 1986, Experiments 1 and 2; see also Nakatani, 2006, for related evidence from Japanese).
- Verb clusters with more than two argument-taking verbs are difficult (Bach et al., 1986, Experiment 1).
- Passivization of the argument structure resulting from combining two argument-taking verbs is difficult (Bader and Schmid, 2009a).

Based on these findings, (15) gives a descriptively more adequate formulation of the Verb-Cluster Complexity Hypothesis of Bader and Schmid (2009a).

- (15) *Verb-Cluster Complexity Hypothesis (revised)*  
Operating on a composite argument structure derived by verb-cluster formation is costly for the human parser.

Combining two argument-taking verbs creates a composite argument structure. This is an easy task for the parser, but applying further operations to such a composite argument structure is difficult. These further operations can be of two types. First, a third argument-taking verb is added, as in the sentences of Bach et al. (1986) and Experiment 1. Second, an argument-structure changing auxiliary is added, as in the long passive construction investigated by Bader and Schmid (2009a). Adding a verb that has no effect on the argument structure of the verb cluster it combines with (e.g., a perfect auxiliary) is easy. In sum, working on simple argument structures as they are associated with verbs is easy for the parser, but working on composite argument structures is difficult. A possible reason for this could be that composite argument structures cannot be retrieved from

the lexicon but must be computed on the fly. The need to hold the resulting complex argument structure in working memory and simultaneously to work on it might be the source of the observed difficulty.

A final issue concerning verb-cluster formation is why Bach et al. (1986) found Dutch crossed dependencies to be somewhat more acceptable than German nested dependencies. As noted above, the size of this effect was rather small, and several minor advantages brought about by the Dutch order could be responsible for it. First, the order of verbs in Dutch is better suited for incremental parsing and interpretation than the order of verbs in German. Consider first Dutch. The crossed dependencies of Dutch are a consequence of the fact that the hierarchically highest verb  $V_1$  comes first, followed by  $V_2$ , that is, the verb selected by  $V_1$ , and so on. Verbs thus appear in the same order as in English. Due to this ordering, Dutch verb clusters can be syntactically analyzed and semantically interpreted incrementally as each verb is encountered. The first verb to be encountered is the finite verb. This verb can be linked to the first NP, the subject NP, and a preliminary semantic analysis can be computed with an open slot for the missing verbal complement. This open slot can be filled on encountering the second verb and the second NP can be linked. There will now be an open slot for the verbal complement of the second verb, which is filled as soon as the third verb is encountered.

Since verbs in German appear in reversed order, parsing and interpretation cannot be fully incremental. When the first verb of the cluster is encountered, the third NP can be linked as its subject argument, but how the verb is related to the already build syntactic structure or to the partial semantic representation computed so far cannot be determined, because this verb is a non-finite verb, but a finite verb is required to make contact with the existing higher level structures. The second verb is again a non-finite verb, so making the connection with the higher level structure has still to wait, although linking of its subject argument is possible. Only when the third verb, the finite verb, is encountered, is it possible to fully integrate the syntactic structure and the semantic representation of the embedded clause into those of the matrix clause.

The processing advantage for crossed dependencies with regard to incremental parsing does not seem to be a large one, but the acceptability difference found by Bach et al. (1986) was not large either. Furthermore, other factors may also contribute to this difference. For example, it has been hypothesized that the order of the arguments associated with a verb reflect their hierarchical position within the semantic representation of the verb (e.g., Bierwisch, 1986). The agent is the highest argument in the semantic representation (as the first argument of the causal relation), and at the same time the argument that precedes all other arguments. Given this hypothesis, a Dutch verb cluster, where the semantically highest verb comes first, is advantageous because the order of verbs parallels the order of arguments.

In the remainder of this paper, sentences with four nested dependencies and verb clusters containing at most two verbs will be explored more closely. Experiment 2 investigates the complexity of sentences for which storage cost and integration cost make opposite predictions. The final Experiment 3 takes a

closer look at integration cost in sentences matched for storage cost.

## 4. EXPERIMENT 2

As noted above, prior research on sentence complexity in verb-final languages has revealed an anti-locality effect: additional material in front of the clause-final verb leads to shorter reading times on the verb (e.g., Konieczny, 2000; Vasishth and Lewis, 2006; Nakatani and Gibson, 2008; Levy and Keller, 2013). Locality effects have also been found, however. Levy and Keller (2013) investigated sentences as in (16), varying whether or not the relative clause contained the adverbial and the dative object.

- (16) Der Mitschüler, der (zur Ahndung) (dem  
the classmate, who.NOM as payback the.DAT  
Sohn) den Fußball versteckt hat, ...  
son the.ACC football hidden has  
'The classmate who hid the football from the son as  
payback ...'

Reading times on the relative clause verb were shorter when the dative object was included but longer when the adverbial phrase was included. Following Konieczny (2000), Levy and Keller (2013) explain this in terms of expectations. When the relative clause contains a dative object in addition to an accusative object, a more specific prediction concerning the upcoming verb is possible, making the integration of the verb easier. An additional adverbial phrase, in contrast, is of no help in predicting the verb. In this case, reading times go up due to the lengthened dependency. Thus, both integration cost and verb-specific expectations seem to affect processing cost in verb-final clauses.

What has not been investigated so far is how integration cost and storage cost jointly affect the acceptability of sentences with multiple center embedding. In contrast to the verb specific expectations manipulated in investigations of the anti-locality effect, storage cost is a measure of the number of expectations that the parser has to retain at each point during the parse. Experiment 2 investigates sentences containing four nested dependencies for which integration cost and storage cost make opposing predictions. One type of sentences is similar to the sentences in the 2×2 condition of Experiment 1. This sentence type is illustrated in (17).

- (17) Ich weiß, dass Peter die Behauptung, dass  
I know that Peter.NOM the.ACC claim that  
der Moderator den Sänger auftreten ließ und  
the.NOM host the.ACC singer perform let and  
dann kündigte, zu entkräften versuchte.  
then resigned to refute tried  
'I know that Peter tried to refute the claim that the host  
let the singer perform and then resigned'

Like the 2×2 sentences of Experiment 1, sentence (17) contains one degree of center embedding. As shown in **Table 4**, both the matrix clause and the embedded clause contain a 2-verb

cluster and thus four nested dependencies, two within the matrix clause and two within the embedded clause. While sentences as in (17) are similar to the 2×2 sentences of Experiment 1 with regard to their basic structure (4 nested dependencies distributed across two separate 2-verb clusters), there are also two differences. First, the center-embedded clause in (17) is not a relative clause but a complement clause. This change was made because complement clauses do not involve traces. Traces are a controversial issue in syntactic theory and theories of the human parser alike. By investigating complement clauses instead of relative clauses, these controversies are circumvented when integration cost profiles are computed. The second difference is that the 2-verb cluster in the center-embedded clause is followed by a conjoined VP. This conjoined VP does not increase the degree of nesting but increases the distance between the verb cluster of the upper *that*-clause and its arguments.

Sentences as in (17) will be compared to sentences as in (18).

- (18) Ich weiß, dass Peter die Behauptung, dass  
I know that Peter.NOM the.ACC claim that  
der Moderator, nachdem der Sänger  
the.NOM host after the.NOM singer  
aufgetreten war, kündigte, zu entkräften versuchte.  
performed has resigned to refute tried  
'I know that Peter tried to refute the claim that the host  
resigned after the singer had performed.'

As shown in **Table 4**, sentence (18) again contains four nested dependencies, but this time distributed across three verb clusters. The upper *that*-clause contains a 2-verb cluster. Center-embedded within the upper *that*-clause is a second *that*-clause which contains a 1-verb cluster. The lower *that*-clause in turn hosts a center-embedded temporal clause which also contains a 1-verb cluster. The degree of center embedding is two in sentence (18).

Storage cost is greater in (18) than in (17), but the reverse is true for integration cost. Consider first storage cost. When processing the most deeply embedded clause in sentence (17), the parser must keep two predicted verbal heads in memory—one verb for each *that*-clause. For sentence (17), one additional predicted verb must be kept in memory, the verb of the temporal clause embedded within the lower *that*-clause. Thus, a maximal storage cost of two for sentence (17) contrasts with a maximal storage cost of three for sentence (18). According to the Predicted Syntactic Head Hypothesis in (6), sentence (18) should therefore be more difficult to process than sentence (17).

Consider next integration cost, which is also shown in **Table 4**. The first line below each sentence gives the referential processing cost (RC), the second line the structural integration cost (IC), and the final line the total processing cost, which is the sum of referential and integration cost. Consider first the integration cost profile for center embedded sentences with one level of embedding. Each NP and each verb introduces a new discourse referent and is therefore associated with a referential cost of 1.

**TABLE 4 |** Syntactic dependencies and integration cost profiles for the sentence conditions of Experiment 2.

CENTER EMBEDDED – 1 EMBEDDING												
	dass	P	die Behauptung,	dass	der M,	den S	auftreten	ließ	und dann	kündigte,	zu entkräften	versuchte.
	that	P	the claim	that	the m	the s	perform	let	and then	resigned	to refute	tried
RC	0	1	1	0	1	1	1	1	0	1	1	1
IC	0	0	0	0	0	0	0	2+0	0	3	5	7+0
Total	0	1	1	0	1	1	1	3	0	4	6	8
CENTER EMBEDDED – 2 EMBEDDINGS												
	dass	P	die Behauptung,	dass	der M,	nachdem	der S	aufgetreten war,	kündigte,	zu entkräften	versuchte.	
	that	P	the claim	that	the m	after	the s	performed has	resigned	to refute	tried	
RC	0	1	1	0	1	0	1	1	1	1	1	
IC	0	0	0	0	0	0	0	0	2	4	6+0	
Total	0	1	1	0	1	0	1	1	3	5	7	
CONTROL – 1 EMBEDDING												
	dass	P	die Behauptung	zu entkräften	versuchte,	dass	der M	den S	auftreten	ließ	und dann	kündigte.
	that	P	the claim	to refute	tried	that	the m	the s	perform	let	and then	resigned
RC	0	1	1	1	1	0	1	1	1	1	0	1
IC	0	0	0	0	2+0	2	0	0	0	2+0	0	3
Total	0	1	1	1	3	2	1	1	1	3	0	4
CONTROL – 2 EMBEDDINGS												
	dass	P	die Behauptung	zu entkräften	versuchte,	dass	der M,	nachdem	der S	aufgetreten war,	kündigte.	
	that	P	the claim	to refute	tried	that	the m	after	the s	performed has	resigned	
RC	0	1	1	1	1	0	1	0	1	1	1	
IC	0	0	0	0	2+0	2	0	0	0	0	2	
Total	0	1	1	1	3	2	1	0	1	1	3	

RC, referential processing cost; IC, structural integration cost; total, total processing cost.

As for structural integration cost, the following considerations apply:

- *Paul, die Behauptung* (“the claim”), *der Moderator* (“the host”), *den Sänger* (“the singer”): There is no head with which the NPs could integrate directly after they have been encountered in the input. Structural integration cost is therefore 0 for all of them.
- *auftreten* (“perform”): The first verb that the parser encounters in the input can be integrated with the final NP *den Sänger*, which fills the subject argument role of *auftreten*. Integration cost is zero because the two items occur adjacent to each other.
- *ließ* (“let”): Integrating *ließ* with its subject *der Moderator* (“the host”) spans two new discourse referents, *den Sänger* (“the singer”) and *auftreten* (“perform”). This verb must also integrate with its verbal complement *auftreten*, but because the two verbs occur adjacent to each other, no structural integration costs ensue.

- *kündigte* (“resigned”): This verb also integrates with the NP *der Moderator* (“the host”). Three new discourse referents intervene between the verb and its subject.
- *zu entkräften* (“refute”): The integration of this verb with its object *die Behauptung* (“the claim”) spans 5 new discourse referents.
- *versuchte* (“tried”): This verb must integrate with its subject and with its verbal complement. The former integration spans 7 new discourse referents whereas the latter spans zero new discourse referents.

Center embedded sentences with two levels of embedding have the same integration cost profile up to and including the first verb in the left-to-right parse. Because the verb *ließ* (“let”) does not occur in 2-embedding sentences, the final three verbs are associated with a smaller integration cost in 2-embedding sentences than in 1-embedding sentences. For example, the verb

*kündigte* (“resigned”) is now the second verb. As before, it must integrate with the NP *der Moderator* (“the host”). This integration spans only two new referents (singer, perform) in contrast to three for 1-embedding sentences (singer, perform, let). For the last two verbs, integration cost is similarly diminished by one unit in 2-embedding sentences.

For center embedded sentences with one level of embedding, the maximum integration cost is 8 whereas for center embedded sentences with two levels of embedding the maximum integration is only 7. If we assume with Gibson (2000) that the acceptability ratings for a sentence reflect its maximum integration cost, we get the prediction that center embedded sentences with two levels of embedding should be more acceptable than center embedded sentences with one level of embedding. The same holds for summed integration cost, which is obtained by summing up the integration cost for each word. Summed integration cost is 26 for center embedded sentences with one level of embedding but only 20 for center embedded sentences with two levels of embedding.

As in Experiment 1, control sentences in Experiment 2 were derived from experimental sentences by means of extraposition. As shown in **Table 4**, the most deeply embedded *that*-clause was put behind the higher *that*-clause. For 1-embedding sentences, this removes center embedding completely, and at no point during the ongoing parse the parser has to keep more than a single predicted verb in memory. For 2-embedding sentences, the extraposed *that*-clause still contains a center-embedded temporal clause. When processing this temporal clause, two predicted verbs must be kept in memory. The maximum storage cost for control sentences is therefore lower than for center-embedded sentences, but the difference between the two types of control sentences is the same as the difference between the two center-embedded sentences (2 vs. 1 for control sentences, 3 vs. 2 for center-embedded sentences). Integration cost is similarly reduced in control sentences in comparison to center-embedded sentences. Furthermore, the control sentences are similar to the center embedded sentences in that integration cost is lower in sentences with two embeddings than in sentences with one embedding. This holds for maximum and summed integration cost alike.

In sum, integration cost is higher in 1-embedding sentences than in 2-embedding sentences, but storage cost is higher in 2-embedding sentences than in 1-embedding sentences. This holds for center embedded and for control sentences alike, although storage and integration cost are lower in the latter than in the former.

## 4.1. Methods

### 4.1.1. Participants

Fourty students from the Goethe-University Frankfurt participated in Experiment 2. All participants were native speakers of German and naive with respect to the purpose of the experiment. Ethical approval was not required for this study in accordance with the national and institutional guidelines.

### 4.1.2. Materials

Sixteen sentences were constructed for Experiment 2, with each sentence appearing in four versions according to the two factors

Embedding (1 vs. 2) and Structure (center embedded vs. control). Most sentences were based on the lexical material of the sentences investigated in Experiment 1. The two verbs *versuchen* (“to try”) and *lassen* (“to let”) were used again as verbs selecting a verbal complement. In order to test whether the acceptability of the sentences under consideration is mainly governed by structural factors, not by lexical factors, the position of *versuchen* (“to try”) and *lassen* (“to let”) was varied as a within-item factor.

All sentences again started with the main clause “Ich weiß” (“I know”), followed by a *that*-clause. All *that*-clauses consisted of a proper name as subject, a definite NP as accusative object and a 2-verb cluster. Eight sentences contained a verb cluster with a non-finite lexical verb and the finite verb *ließ* (“let”). The verb cluster of the other eight sentences contained a non-finite lexical verb and the finite control verb *versuchte* (“tried”). **Table 5** shows an example of each sentence type.

The accusative object in all *that*-clauses was a definite NP with a head noun selecting a *that*-clause itself. In 1-embedding sentences, this second *that*-clause started with the subject, followed by an accusative object and a 2-verb cluster. This cluster consisted of a non-finite lexical verb and either the finite verb *ließ* (“let”) or the finite verb *versuchte* (“tried”). When *ließ* appeared in the inner *that*-clause, *versuchte* appeared in the outer *that*-clause, and vice versa. The 2-verb cluster of the inner *that*-clause was followed by the conjunction *und* (“and”), a one-word adverbial and a finite lexical verb. For control sentences, the lower *that*-clause was extraposed behind the upper *that*-clause.

2-embedding sentences differed from 1-embedding sentences as follows. The lower *that*-clause now consisted only of the former subject and the lexical verb that follows the conjunction in 1-embedding sentences. The accusative object and the lexical verb of the 2-verb cluster in 1-embedding sentences were used to construct an adverbial clause that was center-embedded within the lower *that*-clause. The former accusative object was always the subject in this adverbial clause. Control sentences were again created by extraposing the lower *that*-clause behind the higher *that*-clause.

### 4.1.3. Procedure

As in Experiment 1, participants received a written questionnaire and had to rate the acceptability of each sentence on a scale from 1 (“totally unacceptable”) to 7 (“totally acceptable”).

## 4.2. Results

The statistical analysis proceeded as for Experiment 1. An initial inspection revealed that the order of the two verbs *lassen* (“let”) and *versuchen* (“try”) (see **Table 5**) had no effect on acceptability. In all four combinations of the two factors Embedding and Structure, the difference between the two verb orders was less than 0.3, and verb order as a third factor within the statistical model was not involved in any significant effects. The results of Experiment 1 thus seem to reflect the particular syntactic configurations under investigation and not verb-specific idiosyncrasies. The factor verb order was accordingly dropped from all further analyses.

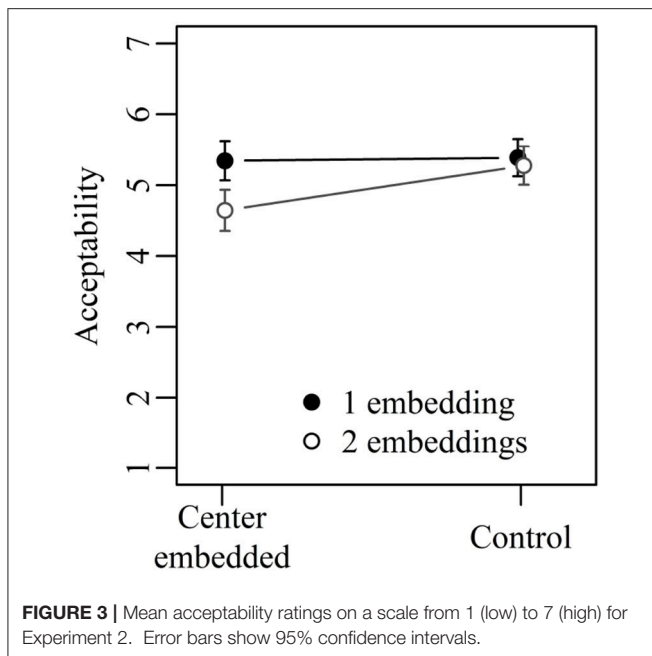
**TABLE 5 |** Two stimulus sentences from Experiment 2, one with a causative verb in the most deeply embedded clause and a control verb in the dominating matrix clause and one with the reversed positions of causative and control verb.

CONTROL VERB IN HIGHER CLAUSE, CAUSATIVE VERB IN LOWER CLAUSE		
1 embedding	Center embedded	dass Moritz das Gerücht, dass der Architekt den Gärtner that M.NOM the.ACC rumor that the.NOM architect the.ACC gardener warten ließ und stattdessen frühstückte, zu verbreiten versuchte. wait let and instead had breakfast to disseminate tried 'that Moritz tried to disseminate the rumor that the architect let the gardener wait and had breakfast instead.'
	Control	dass Moritz das Gerücht zu verbreiten versuchte, dass der Architekt that M.NOM the.ACC rumor to disseminate tried that the.NOM architect den Gärtner warten ließ und stattdessen frühstückte. the.ACC gardener wait let and instead had breakfast
2 embeddings	Center embedded	dass Moritz das Gerücht, dass der Architekt, während der Gärtner that M.NOM the.ACC rumor that the.NOM architect while the.NOM gardener gewartet hat, frühstückte, zu verbreiten versuchte. waited has had breakfast to disseminate tried 'that Moritz tried to disseminate the rumor that the architect had breakfast while the gardener waited.'
	Control	dass Moritz das Gerücht zu verbreiten versuchte, that M.NOM the.ACC rumor to disseminate tried dass der Architekt, während der Gärtner gewartet hat, frühstückte. that the.NOM architect while the.NOM gardener waited had had breakfast
CAUSATIVE VERB IN HIGHER CLAUSE, CONTROL VERB IN LOWER CLAUSE		
1 embeddings	Center embedded	dass der Wirt die Behauptung, dass der Koch den Brand that the.NOM landlord the.ACC claim that the.NOM cook the.ACC fire zu löschen versuchte und dann verzweifelte, verbieten ließ. to extinguish tried and then despaired banned got 'that the landlord let ban the claim that the cook tried to extinguish the fire and despaired.'
	Control	dass der Wirt die Behauptung verbieten ließ, that the.NOM landlord the.ACC claim banned got dass der Koch den Brand zu löschen versuchte und dann verzweifelte. that the cook the fire to extinguish tried and then despaired
2 embeddings	Center embedded	dass der Wirt die Behauptung, dass der Koch, that the.NOM landlord the.ACC claim that the.NOM cook nachdem der Brand gelöscht war, verzweifelte, verbieten ließ. after the.NOM fire extinguished was despaired banned got 'that the landlord let ban the claim that the cook despaired after the fire had been extinguished.'
	Control	dass der Wirt die Behauptung verbieten ließ, that the.NOM landlord the.ACC claim banned got dass der Koch, nachdem der Brand gelöscht war, verzweifelte. that the.NOM cook after the.NOM fire extinguished was despaired

All sentences were introduced by the main clause "Ich weiß" ("I know"). Center-embedded and control sentences have the same meaning and only differ with regard to the position of the embedded clauses. A translation is therefore only given for the center-embedded condition.

Figure 3 shows the mean acceptability ratings for Experiment 2 collapsed across verb order. The results of the corresponding mixed-effect model are given in Table 6. Both the two main effects and the interaction between them were significant. Center embedded sentences with one embedding did not differ significantly from corresponding control sentences (5.3 vs. 5.4; Tukey's test:  $t$ -ratio = 0.22; n.s.). Center-embedded sentences

with two embeddings, in contrast, were judged as significantly less acceptable than corresponding control sentences (4.6 vs. 5.3; Tukey's test:  $t$ -ratio = 3.17;  $p < 0.05$ ). The two types of control sentences did not differ significantly from each other (5.4 vs. 5.3; Tukey's test:  $t$ -ratio = 0.58; n.s.) but the two types of center embedded sentences did (4.6 vs. 5.3; Tukey's test:  $t$ -ratio = 3.63;  $p < 0.01$ ).



**TABLE 6** | Linear mixed model fitted by maximum likelihood estimation for Experiment 2, including *p*-values from likelihood ratio tests.

	Coefficient	Std. Error	t-value	<i>p</i> (LRT)
(Intercept)	5.1625	0.1859	27.769	
Structure	-0.3375	0.1525	-2.213	<0.01
Embeddings	0.4063	0.1435	2.832	<0.01
Structure:embeddings	0.5875	0.2570	2.286	<0.05

*Acceptability* ~ *structure* \* *embedding* + (*structure* + *embedding* || *subject*) + (*structure* \* *embedding* || *sentence*).

### 4.3. Discussion

Experiment 2 investigated two types of center-embedded sentences that were matched in terms of the number of nested dependencies—they contained always four nested dependencies—but differed in terms of storage and integration cost. Structural integration cost was greater in sentences with one embedding than in sentences with two embeddings, whereas storage cost was greater in 2-embedding sentences than in 1-embedding sentences. Since center-embedded sentences with one embedding were judged as more acceptable than center-embedded sentences with two embeddings, Experiment 2 allows the conclusion that storage cost (as measured by the number of predicted heads) is more important than integration cost (as measured by dependency length). In addition, Experiment 2 strengthens the conclusion reached in Experiment 1 that the number of nested dependencies is not a good predictor for sentence complexity. Despite containing four nested dependencies, center-embedded sentences with one embedding were as acceptable as their control sentences, which at no point contained more than two nested dependencies.

The current results show that sentence complexity increases with the number of verbs that have to be predicted. This contrasts with cases where predictions get more specific due to the presence of more arguments, as in the sentences exhibiting the anti-locality effect. For them, more specific predictions decrease complexity according to the most common interpretation of the anti-locality effect (Konieczny, 2000; Vasishth and Lewis, 2006; Levy and Keller, 2013). Taken together, this suggests that predictions help the parser unless too many predictions have to be made simultaneously.

An additional finding of Experiment 2 was that control sentences for both 1- and 2- embedding sentences were judged as equally acceptable, despite showing the same difference in terms of storage cost as the center-embedded sentences. In embedding 1 control sentences, the maximal number of predicted verbs was one whereas it was two in embedding 2 control sentences. Taken together with the results for the center-embedded sentences, we thus see a decrease in acceptability when the number of predicted heads increases from two to three, but not when it increases from one to two.

## 5. EXPERIMENT 3

Four nested dependencies can be realized by three combinations of verbs and verb clusters: two 2-verb clusters, a 2-verb cluster and two single verbs, and four single verbs. The first two configurations were investigated in Experiments 1 and 2. The third experiment investigates the last configuration—each dependency originates in a verb of its own. An example sentence is given in (19). Because there is a verb for each dependency and dependencies are nested, sentence (19) contains three levels of center embedding.

- (19) Der Vorwurf, dass mein Kollege jeden  
 the.NOM charge that my.NOM colleague every.ACC  
 Song, den ein Sänger, den der Chef  
 song that.ACC a.NOM singer that.ACC the.NOM boss  
 nicht kennt, singt, ablehnt, stimmt.  
 not knows sings rejects is-right  
 ‘The charge that my colleague rejects every song that a  
 singer that the boss does not know sings, is true.’

The first aim of Experiment 3 was to test whether sentences with three levels of center embedding lead to clear unacceptability or whether acceptability degrades in a more gradient way. The second aim of Experiment 3 was to test whether integration cost affects the acceptability of sentences that are matched in terms of storage cost. Integration cost is manipulated by varying the number of new discourse referents spanned by the dependencies in complex sentences as in (19). Like all the sentences investigated so far, all NPs in sentence (19) are full NPs with the exception of the relative pronouns. According to the DLT, each full NP introduces a new discourse referent. This distinguishes full NPs from pronominal NPs, which do not introduce new discourse referents. They are therefore not associated with a referential processing cost and they do not count for the computation of structural integration cost. Evidence for this assumption has been

provided by Warren and Gibson (2002), who have shown that English doubly center-embedded relative clauses are easier to comprehend when the most deeply embedded relative clause contains a pronoun instead of a full NP.

Experiment 3 compares sentences like (19) to sentences like (20). Here, two of the full NPs of sentence (19) have been replaced by first-person pronouns. Two NPs were replaced by pronouns in order to increase the chance of observing an effect of integration cost in case such an effect exists.

- (20) Der Vorwurf, dass ich jeden Song,  
 the.NOM charge that I.NOM every.ACC song  
 den ein Sänger, den ich nicht kenne,  
 that.ACC a.NOM singer that.ACC I.NOM not know  
 singt, ablehne, stimmt.  
 sings reject is-right  
 ‘The charge that I reject every song that a singer that I do  
 not know sings, is true.’

**Table 7** shows the integration cost profiles for the sentences investigated in Experiment 3. For each verb, integration cost is higher in the high-load condition, which contains full NPs throughout, than in the low-load condition, in which two full NPs have been replaced by a pronoun. The first verb (*kenne* “know”), for example, must integrate with its subject and its object. The subject is adjacent to the verb and therefore no structural integration cost ensues. The object, that is, the relative pronoun, is separated from the verb by the subject. When the subject is a full NP, structural integration cost is one, but when the subject is a pronoun, structural integration cost is again zero. Similar considerations apply to the remaining verbs. For them, the difference between the high- and low-load condition is always two, either because the verb must integrate with two arguments (*singt* “sings,” *ablehnt* “rejects”) or because two pronouns intervene (*stimmt* “is correct”). Integration cost is highest on the penultimate verb (*ablehnt* “rejects”). In sum, the maximum integration cost is 10 in the high-load condition and 8 in the low-load condition. Summed integration cost, which is obtained by summing up the integration cost for each word, is 31 in the high-load condition and 22 in the low-load condition. Sentences in the low-load condition should therefore be rated as more acceptable than sentences in the high-load condition.

Extrapolation was again used for deriving control sentences from center-embedded sentences, as also shown in **Table 7**. Because of the high degree of center embedding, extrapolation was applied twice for deriving control sentences. This removes center embedding with the exception of the most deeply embedded relative clause, which is still center-embedded in the control sentences. The maximum integration cost for control sentences is 6 in the high-load condition but only 4 in the low-load condition. Thus, maximal integration cost is lower in control than in center-embedded sentences, but the difference between high- and low-load is identical for experimental and for control sentences. Summed integration cost in control sentences is 15 in the high-load and 11 in the low-load condition and thus lower than in center-embedded sentences. In sum, high-load sentences should be less acceptable than low-load sentences,

and center-embedded sentences should be less acceptable than control sentences. This prediction holds for maximum as well as summed integration cost.

In contrast to the main effects of load and structure, the predictions for the interaction between the two factors differ between maximum and summed integration cost. As shown above, for maximum integration cost the difference between low- and high-load condition is 2 for both center-embedded and control sentences. For summed integration cost, the difference between high- and low-load condition is  $15 - 11 = 4$  for the control sentences but  $31 - 22 = 9$  for the center-embedded sentences. Summed integration cost therefore predicts an interaction between load and structure whereas maximum integration cost predicts additive effects. By looking at the interaction, we can thus test the hypothesis of Gibson (2000) that acceptability reflects maximum integration cost and not summed integration cost. This assumption could not be tested in Experiment 2 because there any potential effect of integration cost was offset by an opposite effect of storage cost.

## 5.1. Methods

### 5.1.1. Participants

Fourty students from the Goethe-University Frankfurt participated in Experiment 3. All participants were native speakers of German and naive with respect to the purpose of the experiment. Ethical approval was not required for this study in accordance with the national and institutional guidelines.

### 5.1.2. Materials

For Experiment 3, sixteen new sentences were constructed, with each sentence appearing in four versions according to the two factors Load (low vs. high) and Structure (center embedded vs. control). A sentence in all of its four versions is shown in **Table 8**. Each sentence started with a noun phrase that was the subject of the main clause. The remainder of the main clause made a predication about the subject NP. The head noun of the subject was always a noun taking a sentential complement in the form of a *that*-clause. This clause appeared either adjacent to the head noun (condition center-embedded) or after the main clause (condition control). The *that*-clause consisted of a subject, an accusative object and a verb. The subject was either the first-person pronoun *ich* (“I”) (condition low load) or a full lexical NP (condition high load). The object of the *that*-clause was modified by a relative clause that either appeared directly behind the object (condition center-embedded) or behind the *that*-clause (condition control). In half of the sentences, this relative clause consisted of a subject relative pronoun, a *von* (“by”) PP, and a verb in the passive voice; in the other half of the sentences, the relative clause consisted of an accusative relative pronoun, a subject and an active verb. The second NP in each relative clause was modified by a second relative clause that always appeared adjacent to its head noun. This relative clause was introduced by an accusative relative pronoun in eight sentences and by a relative pronoun contained within a PP in the other eight sentences. The subject of the relative pronoun was either the first-person pronoun *ich* (“I”) (condition low load) or a full lexical NP (condition high load).



**TABLE 7 |** Syntactic dependencies and integration cost profiles for the sentence conditions of Experiment 3.

CENTER EMBEDDED – LOW PROCESSING LOAD (LEXICAL AND PRONOMINAL NPs)													
	Der Vorwurf, the charge	dass that	ich I	jeden Song, every song	den that	ein Sänger, a singer	den who	ich I	nicht not	kenne, know	singt, sings	ablehne, reject	stimmt. is correct
RC	1		0	1	0	1	0	0		1	1	1	1
IC	0		0	0	0	0	0	0		0	1+2	3+4	5
total	1		0	1	0	1	0	0		1	4	8	6
CENTER EMBEDDED – HIGH PROCESSING LOAD (ONLY LEXICAL NPs)													
	Der Vorwurf, the charge	dass that	m. K. m. c.	jeden Song, every song	den that	ein Sänger, a singer	den who	d. C. t. c.	nicht not	kennt, know	singt, sings	ablehnt, reject	stimmt. is correct
RC	1		1	1	0	1	0	1		1	1	1	1
IC	0		0	0	0	0	0	0		0+1	2+3	4+5	7
total	1		1	1	0	1	0	1		2	6	10	8
CONTROL – LOW PROCESSING LOAD (LEXICAL AND PRONOMINAL NPs)													
	Der Vorwurf, the charge	stimmt, is correct	dass that	ich I	jeden Song, every song	ablehne, reject	den that	ein Sänger, a singer	den who	ich I	nicht not	kenne, know	singt. sings
RC	1	1	0	0	1	1	0	1	0	0		1	1
IC	0	0	0	0	0	0+1	1	0	0	0		0+0	1+2
total	1	1	0	0	1	1	1	1	0	0		1	4
CONTROL – HIGH PROCESSING LOAD (ONLY LEXICAL NPs)													
	Der Vorwurf, the charge	stimmt, is correct	dass that	m. K. m. c.	jeden Song, every song	ablehnt, reject	den that	ein Sänger, a singer	den who	d. C. t. c.	nicht not	kennt, know	singt. sings
RC	1	1	0	1	1	1	0	1	0	0		1	1
IC	0	0	0	0	0	0+1	1	0	0	0		0+1	2+3
total	1	1	0	1	1	1	1	1	0	0		2	6

RC, referential processing cost; IC, structural integration cost; total, total processing cost.

**5.1.3. Procedure**

Acceptability was tested using a questionnaire in the same way as in the two preceding experiments.

**5.2. Results**

The data analysis proceeded as in the preceding experiments. **Figure 4** shows the mean acceptability ratings obtained in Experiment 3. The results of the linear mixed model fitted to the data are given in **Table 9**. The two main effects were significant but the interaction between them was not. Low-load sentence were judged as more acceptable than high-load sentences (5.2 vs. 4.4) and control sentences were judged as more acceptable than center-embedded sentences (5.1 vs. 4.5).

**5.3. Discussion**

Experiment 3 has yielded two major results. The first one is that sentences with lower integration cost were more acceptable than sentences with higher integration cost. Thus, when storage cost is held constant, effects of integration cost become visible.

Furthermore, the effect of integration cost on acceptability was equal in size for center-embedded and for control sentences. This supports Gibson’s (2000) hypothesis that acceptability ratings reflect maximum integration cost, since in both the center-embedded and the control condition, high- and low-load sentences differed by the same amount of two memory units when considering maximum integration cost. When considering summed integration cost, in contrast, the integration cost difference for center-embedded sentences was twice as high as for control sentences. This should have resulted in an interaction between structure and load, but no interaction was found.

The second major finding yielded by Experiment 3 is that even sentences with three degrees of center embedding were far from total unacceptability. In the low-load condition, center-embedded sentences received a mean acceptability rating of 4.8, which is well above the midpoint of the 1-to-7 scale. In the high-load condition, mean acceptability was 4.1 for center-embedded sentences, and thus almost exactly at the midpoint. One reason for this relative high acceptability despite three levels of center

**TABLE 8** | A complete stimulus sentence from Experiment 3.

LOW PROCESSING LOAD	
Center embedded	Das Gerücht, dass ich jedes Rezept, das ein Koch, den ich the.NOM rumor that I.NOM every.ACC recipe that.ACC a.NOM chef who.ACC I.NOM aus dem Fernsehen kenne, kreiert, nachkoche, ist frei erfunden. from TV know creates cook is freely fictitious 'The rumor that I cook every recipe that a cook that I know from TV creates is a complete fabrication.'
Control	Das Gerücht ist frei erfunden, dass ich jedes Rezept nachkoche, the.NOM rumor is freely fictitious that I.NOM every.ACC recipe cook das ein Koch, den ich aus dem Fernsehen kenne, kreiert. that.ACC a.NOM chef who.ACC I.NOM from TV know creates 'The rumor is a complete fabrication that I cook every recipe that a cook that I know from TV creates.'
HIGH PROCESSING LOAD	
Center embedded	Das Gerücht, dass der Sohn jedes Rezept, das ein Koch, the.NOM rumor that the.NOM son every.ACC recipe that.ACC a.NOM chef den der Vater aus dem Fernsehen kennt, kreiert, nachkocht, ist frei erfunden. who.ACC the.NOM father from TV know creates cooks is freely fictitious 'The rumor that the son cooks every recipe that a cook that the father knows from TV creates is a complete fabrication.'
Control	Das Gerücht ist frei erfunden, dass der Sohn jedes Rezept nachkocht, the.NOM rumor is freely fictitious that the.NOM son every.NOM recipe cooks das ein Koch, den der Vater aus dem Fernsehen kennt, kreiert. that a.NOM chef who.ACC the.NOM father from TV know creates 'The rumor is a complete fabrication that the son cooks every recipe that a cook that the father knows from TV creates.'

**TABLE 9** | Linear mixed model fitted by maximum likelihood estimation for Experiment 3, including *p*-values from likelihood ratio tests.

	Coefficient	Std. Error	t-value	p (LRT)
(Intercept)	4.8000	0.1503	31.94	
Structure	-0.6875	0.1904	-3.61	<0.001
Load	0.7188	0.1636	4.39	<0.01
Structure:load	-0.1125	0.2882	-0.39	>0.1

*Acceptability* ~ *structure* \* *load* + (*structure* \* *load* || *subject*) + (*structure* \* *load* || *sentence*).

embedding may be that the highest center-embedded clause was a complement clause and not a relative clause. That clause type matters in configurations of multiple center embedding has been shown by Chen et al. (2005), who found that processing is easier when a complement clause contains a relative clause than when a relative clause contains a further relative clause (see also Gibson, 1998).

## 6. GENERAL DISCUSSION: SYNTACTIC DEPENDENCIES AND SENTENCE COMPLEXITY

This paper has presented three acceptability experiments investigating processing complexity in sentences with multiply nested dependencies. Experiment 1 compared sentences

with three nested dependencies all originating in a single 3-verb cluster to sentences with four nested dependencies originating in two separate 2-verb clusters. The sentences with three nested dependencies were found to be substantially less acceptable than the sentences with four nested dependencies. This falsifies De Vries et al.'s (2011) generalization that sentences with three or more nested dependencies are difficult or even impossible to process by the human parsing mechanism.

Experiment 2 and Experiment 3 explored alternative sources of sentence complexity. Experiment 2 investigated sentences for which integration and storage cost lead to opposing predictions. The results of Experiment 2 confirmed the predictions from storage cost and thereby showed that storage cost outweighs integration cost as predictor of sentence complexity. Experiment 3 finally showed that integration cost still has an influence on sentence complexity when comparing sentences of equal storage cost.

The experimental findings are supported by an ongoing corpus study that searched for sentences with complex verb clusters in the deWaC corpus, a large corpus of written internet texts (Baroni et al., 2009). The deWac corpus has been annotated for lemma and part of speech information, but it is not a treebank. It was therefore not possible to retrieve sentences by searching for particular syntactic structures. Instead, the search had to proceed by specifying strings of tokens constrained by lexical information. In order not to miss relevant examples, the search string had to be specified rather loosely, making it necessary to remove irrelevant

**TABLE 10** | Authentic examples from the deWaC corpus (Baroni et al., 2009) with 3 or 4 nested dependencies.

### 3 NESTED DEPENDENCIES: 2 UPPER, 1 LOWER

Die Natur erspart den Wissenschaftlern derartige Reisen, indem sie Bruchstücke von Asteroiden, die aus irgendwelchen Gründen **zerborsten sind**, als Meteoriten zur Erde **herabregnen läßt**.

'Nature spares scientists such journeys because it **lets** debris of asteroids **rain down** as meteorites that **burst**ed for some reason.'

### 3 NESTED DEPENDENCIES: 1 UPPER, 2 LOWER

Auch wenn jeder, der einmal die dramatische Baumasse Manhattans aus dem Meer **wachsen sah**, größte Schwierigkeiten **haben dürfte**, sich an diesem Ort Wälder, Hügel, Wiesen und Marschen vorzustellen.

'Even if anyone who **saw** the massive bulk of buildings of Manhattan **growing out** of the sea **might have** difficulties imagining woods, hills, meadows and marsh at this place.'

### 4 NESTED DEPENDENCIES: 2 UPPER, 2 LOWER

Sie erkannten, dass sie zuerst einmal die Kultur des jeweiligen Landes, das sie **zu missionieren beabsichtigten**, **kennen- und schätzen lernen mussten** und nicht mit einer gewissen europäischen Arroganz die dortigen Gepflogenheiten sofort als "Teufelswerk" ablehnen sollten.

'They recognized that first, they should **try to get to know and appreciate** the culture of the country they **are aiming to evangelize** and that they should not reject local customs as a creation of the devil.'

sentences by hand. For that reason, quantitative information is not yet available for the structures under consideration, although certain tendencies are discernible.

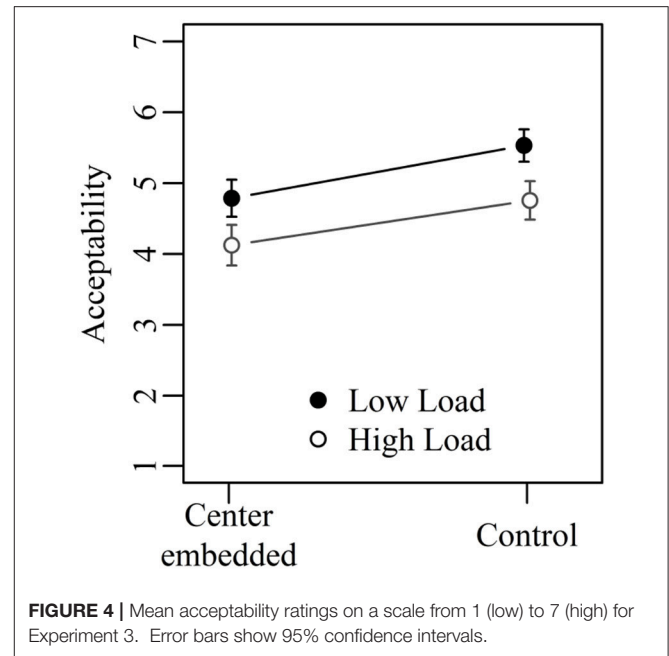
A large number of sentences with verb clusters containing at least three verbs were found, but none of the types investigated by Bach et al. (1986) and in Experiment 1. Instead, complex clusters either contained at most two argument-taking verbs plus additional non-argument-taking verbs (auxiliaries, modals). Sentences with 3 or more nested dependencies distributed across two separate verb clusters were found, however, as shown in **Table 10**. Overall, such examples are rare, but this is not unexpected because they must contain subpatterns that are themselves not very frequent, namely an embedded clause that has not been extraposed, and at least one verb cluster with two argument-taking verbs. Crucially, examples with three or more nested dependencies do occur, and they are not particularly difficult to comprehend, in accordance with the experimental results yielded by the preceding experiments.

Before going on, it should be pointed out that sentences that are easy to comprehend despite containing three or more nested dependencies are nothing special about German. Relevant examples may be somewhat easier to construct in a verb-final language, but they can also be found in an SVO language like English, as shown by the examples in (21) and (22).

(21) I wonder which book the author who I hope to meet at the festival will talk about

(22) It is Mary who the manager that I talked to tried to convince

(21) and (22) both contain only a single level of center embedding, but nevertheless four nested dependencies, two in the upper clause and two in the lower clause. Thus, as in German,



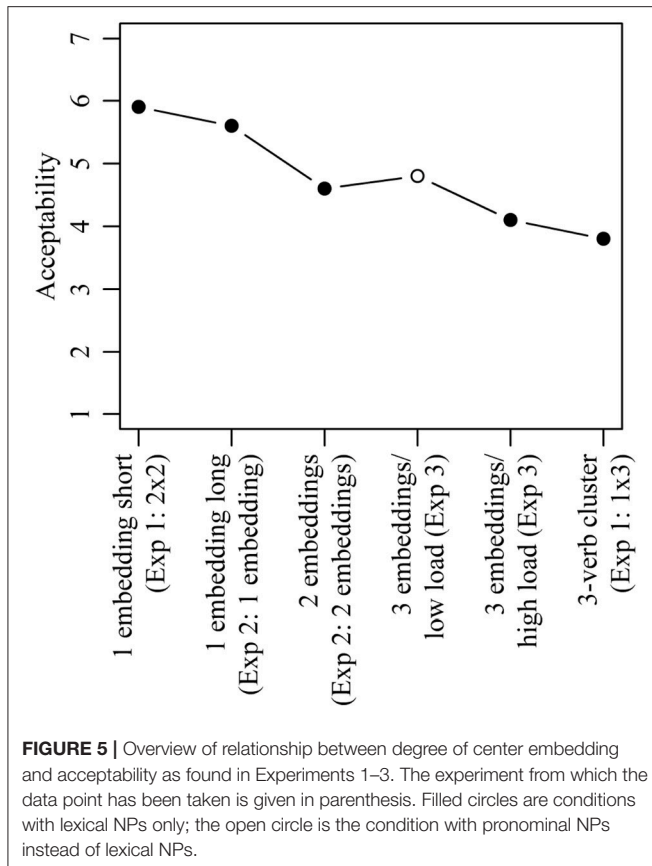
**FIGURE 4** | Mean acceptability ratings on a scale from 1 (low) to 7 (high) for Experiment 3. Error bars show 95% confidence intervals.

increasing the number of nested dependencies without increasing the degree of center embedding makes English sentences not overly hard to process.

The major question pursued in this paper concerned the role of dependency formation in accounting for syntactic complexity. With regard to this question, the results yielded by Experiments 1–3 indicate that the number of nested dependencies that a sentence contains is a poor predictor of sentence complexity. Thus, the *Incomplete Dependency Hypothesis* in (3) is invalid for dependencies independently of their order. In contrast to the number of dependencies, dependency length, as captured in terms of integration cost, was found to affect sentence complexity. Importantly, however, storage cost, a measure not related to dependency formation but to phrase-structure building, turned out to be more important than integration cost, which had an effect only when storage cost was held constant.

Storage cost was measured by the number of predicted verbal heads. Because each clause obligatorily contains a verbal head, storage costs measured in this way directly reflects degree of center embedding. One degree of center embedding is associated with a maximal storage cost of two verbal heads, two degrees of center embedding are associated with a maximal storage cost of three verbal heads, and so on. To explore the

relationship between storage cost/degree of center embedding and acceptability in more detail, **Figure 5** provides a graphical summary of the results obtained in the preceding experiments.



**Figure 5** shows that acceptability decreases with each additional step of center embedding as long as only sentences with full NPs are considered. The decrease in acceptability is modest at each step, and sentences with three degrees of center embedding are still above the midpoint of the 1-to-7 scale. About the same value was observed for sentences with 3-verb clusters as investigated in Experiment 1. We can therefore conclude that at least up to a degree of three center embeddings, acceptability decreases in a gradient fashion. If the trend visible in **Figure 5** continues, sentences with still further levels of center embedding will become more and more unacceptable. A further finding visible in **Figure 5** is that sentences with three degrees of center embedding and pronominal NPs in place of some of the full NPs were as acceptable as sentences with only two degrees of center embedding but with full NPs throughout. In sum, the degree of center embedding seems to be an important, or perhaps even the

## REFERENCES

- Bach, E., Brown, C., and Marslen-Wilson, W. D. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Lang. Cogn. Process.* 1, 249–262. doi: 10.1080/01690968608404677
- Bader, M., and Schmid, T. (2009a). Minimality in verb-cluster formation. *Lingua* 119, 1458–1481. doi: 10.1016/j.lingua.2008.03.003

most important, predictor of acceptability, but its influence can be modulated by other factors.

The finding that acceptability degrades gracefully with increasing degree of center embedding makes it unlikely that a categorical limit on center embedding can be found, a limit that cleanly separates acceptable from unacceptable embedding. This argues against theories that ascribe the severe limitation on center embedding to the existence of a memory system that provides only a small, fixed amount of storage space for processing center embedded sentences (e.g., Yngve, 1960; Kimball, 1975; Stabler, 1994). Instead, this graceful degradation argues in favor of a multi-factorial account of the limits on center embedding. Two factors affecting sentence complexity are storage cost and integration cost, as shown by the experiments reported above. Other general factors that have been invoked to explain sentence complexity are frequency (e.g., Hale, 2011) and interference (e.g., Van Dyke and McElree, 2006; Belletti and Rizzi, 2013). Furthermore, for the case of sentences with multiple center embedding, Fodor (2013) has proposed that parsing difficulties arise because of difficulties with assigning a prosodic structure to such sentences. For reasons of space, it must be left as a task for future research to determine how the complexity of syntactic parsing follows from the joint work of the various factors.

To conclude, the results reported in this paper add to the existing evidence that the sheer number of open dependencies is not a crucial factor determining sentence complexity, independently of the the order of the dependencies. It is true that in many cases, more complex sentences contain more nested dependencies, but such sentences are typically also more complex in other ways. For example, sentences with doubly center-embedded relative clauses usually contain more nested dependencies than sentences with only a single degree of embedding, but they are also more complex in terms of storage cost, for example.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02268/full#supplementary-material>

- Bader, M., and Schmid, T. (2009b). Verb clusters in Colloquial German. *J. Comp. German. Linguist.* 12, 175–228. doi: 10.1007/s10828-009-9032-x
- Bader, M., Schmid, T., and Häussler, J. (2009). “Optionality in verb-cluster formation,” in *The Fruits of Empirical Linguistics. Vol. 2: Product*, eds S. Winkler, and S. Featherston (Berlin:de Gruyter), 37–58.
- Baroni, M., Bernardini, S., Ferraresi, A., and Zanchetta, E. (2009). The WaCky Wide Web: A collection of very large linguistically processed web-crawled corpora. *Lang. Resour. Eval. J.* 23, 209–226. doi: 10.1007/s10579-009-9081-4

- Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015a). Parsimonious mixed models. *arXiv:1506.04967* [stat.ME].
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Belletti, A., and Rizzi, L. (2013). "Intervention in grammar and processing," in *From Grammar to Meaning: The Spontaneous Logicality of Language*, eds I. Caponigro and C. Cecchetto (Cambridge: University Press, Cambridge), 293–311.
- Bierwisch, M. (1986). "On the nature of semantic form in natural language," in *Human Memory and Cognitive Capabilities. Mechanisms and Performances*, eds F. Klix and H. Hangendorf (Amsterdam: North Holland), 765–784.
- Chen, E., Gibson, E., and Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *J. Mem. Lang.* 52, 144–169. doi: 10.1016/j.jml.2004.10.001
- De Vries, M., Christiansen, M. H., and Petersson, K. M. (2011). Learning recursion: multiple nested and crossed dependencies. *Biolinguistics* 5, 010–035.
- Fodor, J. (2013). "Pronouncing and comprehending center-embedded sentences," in *Language Down the Garden-Path: The Biological and Cognitive Basis for Linguistic Structure*, eds M. Sanz, I. Laka, and M. K. Tanenhaus (Oxford: Oxford University Press), 206–228.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1–76. doi: 10.1016/S0010-0277(98)00034-1
- Gibson, E. (2000). "The dependency locality theory: a distance-based theory of linguistic complexity," in *Image, Language, Brain. Papers from the First Mind Articulation Project Symposium*, eds A. Marantz, Y. Miyashita, and W. O'Neil (Cambridge, MA: MIT Press), 95–126.
- Gildea, D., and Temperley, D. (2010). Do grammars minimize dependency length? *Cogn. Sci.* 34, 286–310.
- Hale, J. T. (2011). What a rational parser would do. *Cogn. Sci.* 35, 399–443. doi: 10.1111/j.1551-6709.2010.01145.x
- Hawkins, J. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Hawkins, J. A. (2014). *Cross-Linguistic Variation and Efficiency*. Oxford: Oxford University Press.
- Jaeger, T. F., and Tily, H. (2011). On language 'utility': Processing complexity and communicative efficiency. *Wiley Interdiscipl. Rev. Cogn. Sci.* 2, 323–335. doi: 10.1002/wcs.126
- Kimball, J. (1975). "Predictive analysis and over-the-top parsing," in *Syntax and Semantics, Vol. 4*, ed J. Kimball (New York, NY: Academic Press), 155–179.
- Konieczny, L. (2000). Locality and parsing complexity. *J. Psycholinguist. Res.* 29, 627–645. doi: 10.1023/A:1026528912821
- Levy, R. (2013). "Memory and surprisal in human sentence comprehension," in *Sentence Processing*, ed R. G. van Gompel (New York, NY: Psychology Press), 78–114.
- Levy, R. P., and Keller, F. (2013). Expectation and locality effects in German verb-final structures. *J. Mem. Lang.* 68, 199–222. doi: 10.1016/j.jml.2012.02.005
- Nakatani, K. (2006). Processing complexity of complex predicates: a case study in Japanese. *Linguist. Inquiry* 37, 625–647. doi: 10.1162/ling.2006.37.4.625
- Nakatani, K., and Gibson, E. (2008). Distinguishing theories of syntactic expectation cost in sentence comprehension: evidence from Japanese. *Linguistics* 46, 63–87. doi: 10.1515/LING.2008.003
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Schmid, T., Bader, M., and Bayer, J. (2005). "Coherence - an experimental approach," in *Linguistic Evidence*, eds M. Reis and S. Kepsers (Berlin: de Gruyter), 435–456.
- Schütze, C. T., and Sprouse, J. (2014). "Judgment data," in *Research Methods in Linguistics*, eds R. J. Podesva and D. Sharma (Cambridge: Cambridge University Press), 27–50.
- Seuren, P. A. M., and Kempen, G., (Eds.). (2003). *Verb Constructions in German and Dutch*. Amsterdam: John Benjamins. Current Issues in Linguistic Theory 242.
- Stabler, E. P. (1994). "The finite connectivity of linguistic structure," in *Perspectives on Sentence Processing*, eds J. Charles, L. Frazier, and K. Rayner (Hillsdale, NJ: Lawrence Erlbaum), 303–336.
- Tallerman, M., Newmeyer, F., Bickerton, D., Bouchard, D., Kaan, E., and Rizzi, L. (2009). "What kind of syntactic phenomena must biologists, neurobiologists, and computer scientists try to explain and replicate?" in *Biological Foundations and Origin of Syntax*, eds D. Bickerton and E. Szathmáry (Cambridge, MA: MIT Press), 135–160.
- Temperley, D. (2007). Minimization of dependency length in written English. *Cognition* 105, 300–333. doi: 10.1016/j.cognition.2006.09.011
- Van Dyke, J. A., and McElree, B. (2006). Retrieval interference in sentence comprehension. *J. Mem. Lang.* 55, 157–166. doi: 10.1016/j.jml.2006.03.007
- Vasishth, S., and Lewis, R. (2006). Argument-head distance and processing complexity: explaining both locality and antilocality effects. *Language* 82, 767–794. doi: 10.1353/lan.2006.0236
- Warren, T., and Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition* 85, 79–112. doi: 10.1016/S0010-0277(02)00087-2
- Weskott, T., and Fanselow, G. (2011). On the informativity of different measures of linguistic acceptability. *Language* 87, 249–273. doi: 10.1353/lan.2011.0041
- Wurmbrand, S. (2006). "Verb clusters, verb raising, and restructuring," in *The Blackwell Companion to Syntax*, eds M. Everaert and H. van Riemsdijk (Oxford: Blackwell), 229–343.
- Wurmbrand, S. (2017). "Verb clusters, verb raising, and restructuring," in *The Blackwell Companion to Syntax*, Vol. 5, eds M. Everaert and H. van Riemsdijk (Oxford: Blackwell), 1–109.
- Yngve, V. H. A. (1960). A model and a hypothesis for language structure. *Proc. Am. Philos. Soc.* 104, 444–466.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Bader. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Scalar and Ignorance Inferences Are Both Computed Immediately upon Encountering the Sentential Connective: The Online Processing of Sentences with Disjunction Using the Visual World Paradigm

Likan Zhan\*

Institute for Speech Pathology and the Brain Science, School of Communication Science, Beijing Language and Culture University, Beijing, China

## OPEN ACCESS

### Edited by:

Aritz Irurtzun,  
Centre National de la Recherche  
Scientifique, France

### Reviewed by:

Lyn Tieu,  
Macquarie University, Australia  
Anne Colette Reboul,  
Claude Bernard University Lyon 1,  
France

### \*Correspondence:

Likan Zhan  
zhanlikan@hotmail.com

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 06 May 2017

**Accepted:** 15 January 2018

**Published:** 31 January 2018

### Citation:

Zhan L (2018) Scalar and Ignorance Inferences Are Both Computed Immediately upon Encountering the Sentential Connective: The Online Processing of Sentences with Disjunction Using the Visual World Paradigm. *Front. Psychol.* 9:61. doi: 10.3389/fpsyg.2018.00061

Accounts based on the pragmatic maxim of quantity make different predictions about the computation of scalar versus ignorance inferences. These different predictions are evaluated in two eye-tracking experiments using a visual world paradigm to assess the on-line computation of inferences. The test sentences contained disjunction phrases, which engender both kinds of inferences. The first experiment documented that both inferences are computed immediately upon encountering the disjunctive connective, at nearly identical temporal locations. The second experiment was designed to determine whether or not there exists an intermediate stage at which the truth of the corresponding conjunction phrase is ignored. No such stage was found.

**Keywords:** scalar implicatures, ignorance inferences, disjunction, grammatical processes, visual-world paradigm

## INTRODUCTION

An utterance in ordinary conversation often expresses information that is stronger than its literal meaning (Grice, 1975). Among such utterances are disjunctions such as [1a]. Literally, the disjunction [1a] is true when at least one of the two disjuncts [2a, 2b] is true. When the two disjuncts [2a, 2b] are both true, the corresponding conjunction [1b] is also true. In ordinary conversation, however, hearing the disjunction [1a] often makes the hearer infer that the corresponding conjunction [1b] is false (*scalar implicature*); and infer that the speaker doesn't know whether either of the two disjuncts [2a, 2b] is true or false (*ignorance inference*). The two inferences result in that the disjunction's actual interpretation is stronger than its literal meaning. It is widely accepted that the two inferences are both generated from a disjunction, but accounts differ in whether they are pragmatic or grammatical.

### 1. The two complex statements

- a. *John's box contains a cow or a rooster.*
- b. *John's box contains a cow and a rooster.*

### 2. The two disjuncts

- a. *John's box contains a cow.*
- b. *John's box contains a rooster.*

First, the pragmatic account contends that both inferences are derived from some post-compositional, pragmatic processes. This account was pioneered by Grice (1975). According to Grice (1975), a speaker sticking to the cooperative principle should be as informative as necessary (maxim of quantity) and should say only things he or she believes to be true (maxim of quality). Hence, a cooperative speaker should assert the strongest statement that he or she is in a position to make. Hearing the speaker's assertion, a hearer then infers that the speaker was not in a position to assert any of the alternative statements that are stronger than the speaker's assertion. The alternative statements constructed from the disjunction [1a] are comprised of two subsets: the first subset is constructed from the Horn scale (Horn, 1972), i.e., [1a, 1b], and the second subset is constructed from the constituents of the disjunction, i.e., [1a, 2a, 2b]. All the relevant alternatives [1b, 2a, 2b] generated from the disjunction are stronger than the disjunctive statement [1a]. Hence, upon hearing the disjunctive statement [1a], a hearer infers that the speaker was in a position to determine the disjunction [1a] is true (maxim of quality), but was not in a position to determine any of the three alternatives [1b, 2a, 2b] is true, otherwise the cooperative speaker would have used these alternative(s) (maxim of quantity). The hearer therefore must compute a *primary inference* that the speaker was not in a position to assert that the three alternatives [1b, 2a, 2b] are true. The hearer then needs to judge whether or not the speaker is likely to have an opinion about the truth of the alternatives; this is called the *competence assumption* (Sauerland, 2004; van Rooij and Schulz, 2004). If the hearer makes the competence assumption, he or she infers that the alternatives are false. This process gives rise to *scalar implicatures*. If the hearer does not make the competence assumption, he or she infers that the speaker is ignorant about the truth-value of the alternatives. This leads to an *ignorance inference*. To derive the two required inferences of the disjunction, one needs to hypothesize that the speaker is opinioned on the alternatives derived from the Horn scale, i.e., [1a, 1b], but is not opinioned on the alternatives derived from the constituents of the disjunction [1a, 2a, 2b], resulting in the *scalar implicature* denying [1b] and the *ignorance inference* relative to [2a, 2b].

Second, the hybrid account contends that scalar implicatures are derived using a compositional, grammatical process, whereas ignorance inferences are derived by a post-compositional pragmatic process. On the hybrid account, interpreting a statement begins with hearers determining whether or not they parse the speaker's assertion using a phonologically null "exhaustification" operator (Fox, 2007). If the statement is parsed without this operator, the result is the literal meaning. If the statement is parsed with this operator, the result is an interpretation that includes scalar implicatures and strengthened meanings. Like the pragmatic account, the hybrid account begins the derivation of inferences by establishing a set of relevant alternatives that the speakers might have used in place of the assertion that the speaker made. The covert exhaustification operator is then applied to both the asserted statement and the relevant alternatives. The exhaustification operator is similar in meaning to the focus operator *only*. The output of the application

of the exhaustification operator is a conjunction of propositions. One proposition is that the asserted statement is true. Another proposition is that all relevant alternatives that are not entailed by the asserted statement are false. According to the hybrid account, the exhaustification operator is part of the on-line composition of meaning, rather than post-compositional as on the pragmatic account. Hence, scalar implicatures should be observed on-line, at that point in sentence processing when the lexical item that triggers the exhaustification operator is encountered. On the hybrid account, this can happen when the lexical item is in the middle of a sentence, or at the end. In terms of the disjunction, a scalar implicature should be observed at that point in sentence processing when the sentential connective *or* is encountered, because the sentential connective *or* is the lexical item that triggers the exhaustification operation. With regards to the ignorance inference, by contrast, Fox (2007, 2014) maintains that this inference is derived from maxims of conversation, as on the pragmatic account. On the other hand, Chierchia (2004, 2017) contends that the ignorance inference results from the contradiction that is generated in computing the scalar implicature. As discussed earlier, the exhaustification operator that is triggered by disjunction [1a] can apply either to the Horn scales [1a, 1b], or to the domain alternatives [1a, 2a, 2b]. If the former, then no contradiction is generated; this yields the scalar implicature. If the latter, a contradiction is generated. In this case, the application of the exhaustification operator yields a meaning according to which the disjunction [1a] is true, and the two contradictory disjuncts [2a, 2b] are both false. When a contradiction is derived, the hearer arrives at the ignorance inference.

Third, the radically grammaticalized account put forward by Meyer (2013) contends that both inferences are derived inside the grammatical system of the language apparatus, rather than in the pragmatics. According to this account, an asserted proposition *S* always covertly attaches to an epistemic operator, *K*. Asserting the statement *S*, then, amounts to the assertion *K(S)*, i.e., the speaker knows or believes *S*. When a statement is parsed with the exhaustification operator, *EXH*, the exhaustification operator can apply either above or below the epistemic operator, leading to two legitimate readings, *EXH-K(S)*, *EXH-K-EXH(S)*. When the statement includes the disjunction connective, as in [1a], both readings give rise to the ignorance inference, based on the domain alternatives [1a, 2a, 2b]. However, the two readings yield different inferences when the alternatives include the Horn scales [1a, 1b]. The first reading results into a weaker inference (or primary inference) than the corresponding conjunction, i.e., the speaker is not in a position to know that the corresponding statement with conjunction [1b] is true, resulting in an ignorance inference relative to the conjunction [1b]. The second reading results in a scalar implicature, i.e., the speaker is in a position to know that the corresponding conjunction [1b] is false.

The different predictions made by the three accounts are summarized as **Table 1**. First, the pragmatic account regards the ignorance inference as being triggered by the conversational maxims, and as being the output of a domain general reasoning procedure. As it applies at the level of speech acts, the ignorance inference has to be post-compositionally processed and should

**TABLE 1** | Predictions made by different accounts.

Account	Scalar implicature vs. Ignorance inference	Primary inference
Pragmatic	Scalar implicature > Ignorance inference	Yes
Hybrid	Scalar implicature < Ignorance inference	No
Grammatical		
EXH-K(S)	Scalar implicature = Ignorance inference	Yes
EXH-K-EXH(S)	Scalar implicature = Ignorance inference	No

not be observed until the offset of the test sentences (Sauerland, 2012). The scalar implicature emerges from the ignorance inference to the Horn scales (primary inference), together with the hearer's competence assumption about the speaker. There should exist an intermediate step where participants are ignorant about the truth-value of the corresponding conjunction, and the scalar implicature should occur temporally later than that of the ignorance inference (Chemla and Singh, 2014a,b). Second, the hybrid account deems the scalar implicature as being triggered by a covert lexical operator. Because the output of a domain specific computation takes place within the linguistic system, the scalar implicature is compositional and could be observed prior to the offset of the test sentences. Researchers who advocate the hybrid account differ in the exact mechanisms of how the ignorance inference is derived (Fox, 2007, 2014; Chierchia, 2017), but the general view is that the ignorance inference is derived within the pragmatic component of the language apparatus. Because pragmatic processes normally follow grammatical processes, the ignorance inference is predicted to occur later than scalar implicatures (Chemla and Singh, 2014a,b). As scalar implicatures are derived in a single step, there should be no stage in processing at which an ignorance inference (or primary inference) is applied to the corresponding conjunction. Third, the radical grammatical account (Meyer, 2013) regards both inferences as being derived from the lexical compositional system. Both the scalar implicature and the ignorance inference should arise prior to the offset of the test sentences and should occur almost at the same time. The two supposed legitimate readings differ in whether the primary inference or ignorance inference on the corresponding conjunction should occur. The first reading *EXH-K(S)* ends up in a weak implicature where the truth-value of the corresponding conjunction is ignored. The second reading *EXH-K-EXH(S)* ends up in a scalar implicature where the corresponding conjunction is negated.

To adjudicate between the three accounts (mainly the pragmatic account and the hybrid account), roughly two clusters of studies have been conducted in literature. The first cluster explored whether the scalar implicature is computed locally or globally. If it is a post-compositional operation at the level of speech acts, then the scalar implicature should only be globally computed and could not be locally computed. If it is a compositional lexical operation, then the scalar implicature should also possible be locally computed. There exist two ways to define global and local: whether the scalar implicature can be incrementally observed, i.e., prior to the offset of the test sentences; or whether the final comprehension of a complex statement is based on the scalar implicature applied on the main clause or applied on the subordinate clause. Using visual world

paradigm, researchers have found that the scalar implicature can be incrementally processed, i.e., prior to the offset of the test audios (Breheny et al., 2006, 2013; Grodner et al., 2010; Degen and Tanenhaus, 2015; Foppolo and Marelli, 2017), although under certain experimental settings the processing could be delayed (Huang and Snedeker, 2009). Using complex statements where the scalar quantifiers are embedded under other words such as the universal quantifier *each*, researchers have found that both the global reading (Geurts and Pouscoulous, 2009) and the local reading (Clifton and Dube, 2010; Chemla and Spector, 2011; Chemla et al., 2017) are possible to be constructed. The first cluster of studies seems support the hybrid account. But the second cluster of studies give a different answer. The second cluster explored whether the scalar implicature is a domain specific, automatic process or a domain general, controlled process. If it is a compositional lexical operation, then the scalar implicature should be domain specific, and should be automatically triggered by the scalar quantifiers, regardless of other cognitive processes. The strengthened meaning should be the default meaning, even though it is more complex than the literal meaning. On the contrary, if it is a post-compositional operation at the level of speech acts, the scalar implicature should be a domain general process and should be constrained by other cognitive processes, such as memory. The strengthened meaning should be more difficult to access. Previous literature have found that both the participants' epistemic status (Bergen and Grodner, 2012) and their available working memory sources (Marty and Chemla, 2013) affected the way a scalar implicature is computed. Furthermore, accessing the strengthened meaning is more time consuming (Bott and Noveck, 2004). So the second cluster of studies seem support the pragmatic account.

To summarize, previous studies mainly focused on the scalar implicature entailed by the existential quantifier *some*, i.e., *some but not all* (but see, Chevallier et al., 2008, on disjunctions). No clear-cut evidence so far has been observed to favor one account over another. Furthermore, the ignorance inference engendered by the disjunction has not been experimentally tested in literature.

To recap, accounts differ in the temporal sequences in which the scalar implicature and ignorance inference are computed online. Accounts also differ in whether or not a disjunction (temporally or permanently) triggers an ignorance inference to the corresponding conjunctive statement. To explore these two questions, we reported two eye-tracking studies using the visual world paradigm (Cooper, 1974; Tanenhaus et al., 1995; Zhan et al., 2015). Experiment 1 explored the temporal sequences of the scalar implicature and ignorance inference. Experiment 2 explored the problem of the ignorance inference applied to the corresponding conjunction.

## EXPERIMENT 1

### Method Participants

Thirty-seven postgraduate students from the Beijing Language and Culture University participated in this experiment. All the participants were native speakers of Mandarin Chinese, with



normal or corrected normal visions. They were paid 30CNY (approximately \$5) for their participation.

## Stimuli

A test image involved two animals and four boxes situated at the four quadrants (**Figure 1**). Two properties of the boxes were manipulated: the size and the closeness of a box. The size of a box influenced the animals included in the box, but not participants' epistemic status on that box. A big box always contained two different animals, while a small box always and only contained one animal, no matter whether the box was closed or not. The closeness of a box influenced participants' epistemic status on that box, but not the animals contained in that box. If a box was open, both the speaker and the hearer were in a position to know what animal(s) were contained in that box. If a box was closed, both the speaker and the hearer were not in a position to know what animal(s) were contained in that box.

The test image in Experiment 1 (**Figure 1**, left) consisted of one big box and three small boxes. In the given example, the two animals were a cow and a rooster. The big box *A* was open and contained both a cow and a rooster. Two of the three small boxes *C* and *D* were also open and contained a rooster and a cow, respectively. The third small box *B* was closed. Henceforth, participants were unable to know which animal was in box *B*. But the size of box *B* is small, so participants knew that the small box *B* contained only one animal: it was either a cow or a rooster, but not both. Sixty images like the left panel of **Figure 1** were constructed, with the spatial locations of the four boxes being counterbalanced and with the two involved animals also being changed across images.

Three test sentences (**Figure 2**) were constructed to each test image: A conjunctive statement (**Figure 2A**), a *but*-statement (**Figure 2B**), and a disjunctive statement (**Figure 2C**). One more statement in the form of *Xiaoming's box doesn't contain a rooster but a cow* was used as a filler and was not analyzed in our studies. In each statement, one animal such as the cow in our example was mentioned as the object of the first proposition, while the other such as the rooster was mentioned as the object of the second proposition, respectively. Participants were told that one of the four boxes belongs to *XiaoMing*. Participants' task was to find *XiaoMing's* box according to the test sentence they heard, and press the corresponding button. Participants' online eye-movements on the four boxes as they were listening to the test audios, as well as the boxes participants behaviorally chose, were recorded and used to deduce how the scalar implicature and ignorance inference were processed. The 240 test sentences were then divided into four groups, with each group containing 15 conjunctions, 15 disjunctions, 15 *but*-statements, and 15 filler statements. Each participant saw all the 60 images and heard only one group of the test audios.

The test sentences were recorded by a female native speaker of Mandarin. To make them the same in length and consistent in intonation, all the test audios were exactly the same except for the objects of the two merged disjuncts/conjuncts. To achieve this, we first recorded four example statements, such as (A–C) in **Figure 2**, as well as all the objects that were going to be used in our studies, such as *pig* and *horse*. We then replaced the two objects in the

example statements, i.e., *cow* and *rooster*, with each pair of the recorded objects, such as *pig* and *horse*, resulting in the full list of our test audios. We did a pilot test by asking several native Mandarin-speakers to judge the naturalness of the test sentences in Mandarin, all the interviewees judged the test sentences to be natural Mandarin sentences. The length of the test audios is marked on **Figure 2**.

Given our experimental design, all boxes would be possible candidates, unless a box was ruled out by the computed inference. First, if the ignorance inference to the domain alternatives is engendered from the disjunctive statement [1a], then participants should be in a position not to know the truth values of the two domain alternatives [2a, 2b]. All situations where participants know the truth of the domain alternatives will be ruled out. A small open box means that participants know the truth value of a corresponding domain alternative. Henceforth, computing the ignorance inference will rule out all the small open boxes and result in significant fewer fixations on these boxes. Second, if the scalar implicature and the Horn scales are derived from the disjunctive statement [1a], then participants should be in a position to know the corresponding conjunctive statement [1b] is false. All the situations where the participants are not in a position to know the corresponding conjunctive statement [1b] is false will be ruled out. In this case, the excluded situations consist of not only the situations where the speaker knows that the corresponding conjunctive statement [1b] is true, but also the situations where the speaker is ignorant about the truth of corresponding conjunctive statement [1b]. In our experimental design, a big open box means that participants know the conjunction [1b] is true. A big closed box means that participants don't know whether the conjunction [1b] is true or false. Computing the scalar implicature will rule out all the big boxes and lead to significant few fixations on these boxes, regardless of whether they are open or closed. Third, if the inference derived from the Horn scales is an ignorance inference but not a scalar implicature, then participants will not derive the inference that the conjunction [1b] is false. In this case, only the situations where participants know that the corresponding conjunction [1b] is true will be ruled out, not the situations where participants don't know the truth of the conjunctive statement. This so-called weak inference will rule out the big open boxes, but not the big closed ones, resulting in significant fewer fixations only on the big open boxes.

## Procedure

Participants were seated approximately 64 cm from a 21 inch, 4:3 color monitor with 1,024\*768 pixel resolution. Twenty-seven pixels equaled approximately to 1° of visual angle. The sampling rate of the EyeLink 1000plus eye-tracker was 500 Hz. Viewing was binocular, but only the participant's dominant eye was tracked. The auditory stimuli were presented via a pair of external speakers situated to the two sides of the monitor. At the beginning of the experiment, participants first saw an introduction of the experiment in Mandarin on the screen. The instruction briefly explained the experimental procedure as we described below.

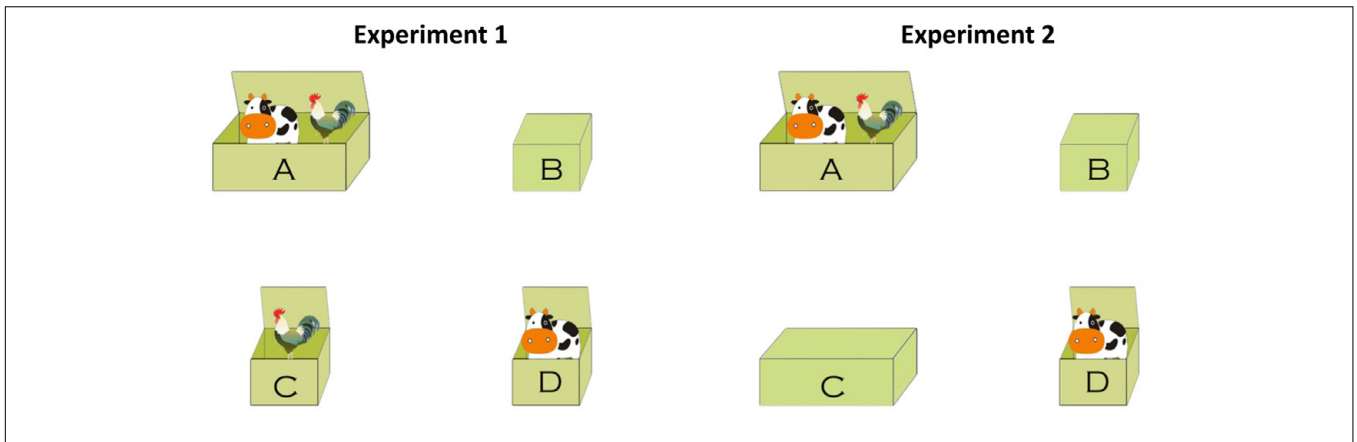


FIGURE 1 | An example of the test images used in Experiment 1 (left) and Experiment 2 (right).

**A And**

小明的 箱子里 有 一只 奶牛 和 一只 公鸡  
 Xiaoming de xiang zi li you yi zhi nai niu he yi zhi gong ji  
 Xiaoming's box in have one-CL cow and one-CL rooster  
*Xiaoming's box contains a cow and a rooster.*

**B But**

小明的 箱子里 有 一只 奶牛 但 没有 公鸡  
 Xiaoming de xiang zi li you yi zhi nai niu dan meiyou gong ji  
 Xiaoming's box in have one-CL cow but not rooster  
*Xiaoming's box contains a cow but not a rooster.*

**C Or**

小明的 箱子里 有 一只 奶牛 或 一只 公鸡  
 Xiaoming de xiang zi li you yi zhi nainiu huo youzhi gongji  
 Xiaoming's box in have one-CL cow or one-CL rooster  
*Xiaoming's box contains a cow or a rooster.*

0s | 0.5s | 1.8s | 2.3s | 3.9s | 5.2s | 6.2s | 7.6s | 8.6s | 9.6s | 11s

0.5s | 1.8s | 1.6s | 1.3s | 1s | 1.4s | 1s | 1s | 1.4s

FIGURE 2 | An example of the test sentences used in Experiment 1 and Experiment 2.

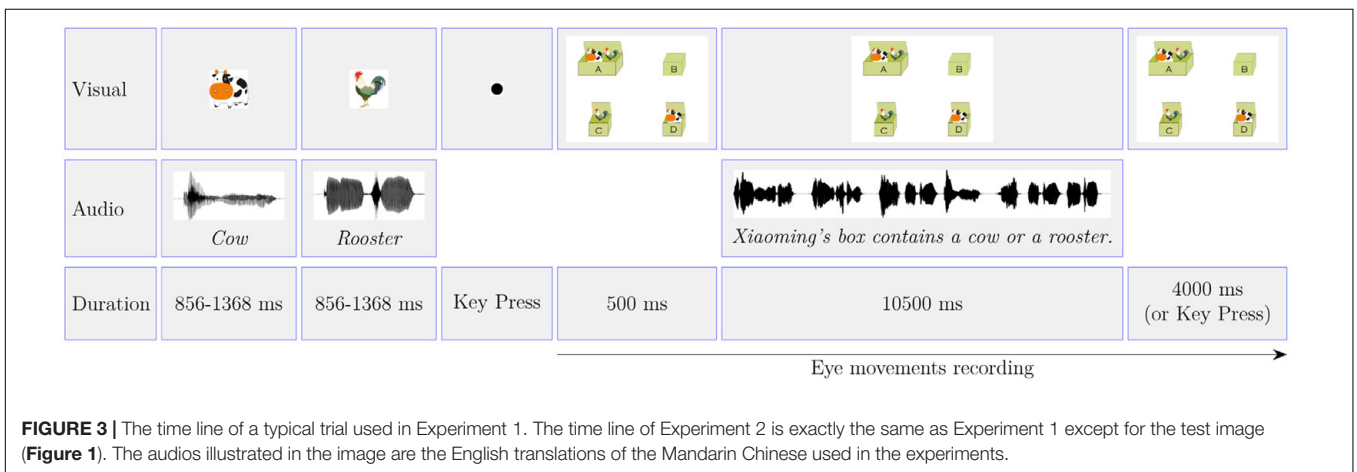
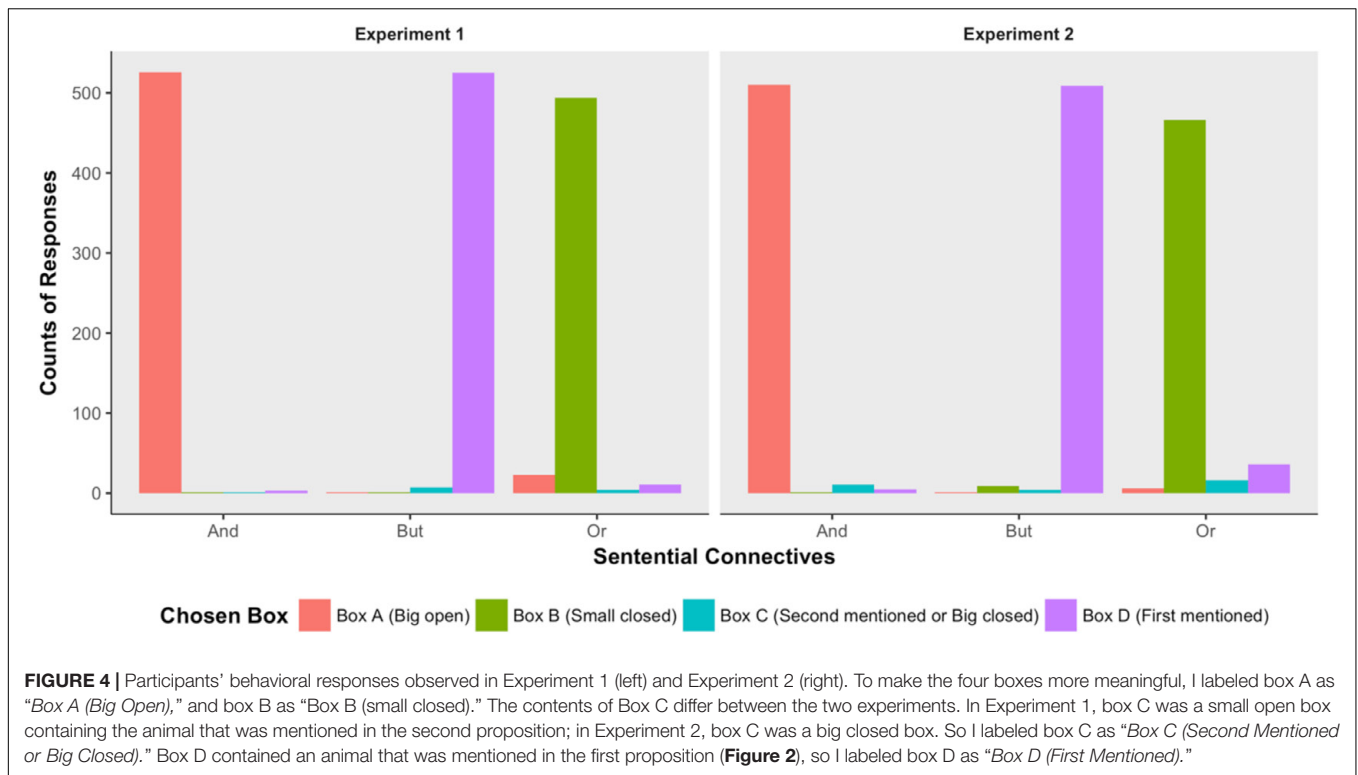


FIGURE 3 | The time line of a typical trial used in Experiment 1. The time line of Experiment 2 is exactly the same as Experiment 1 except for the test image (Figure 1). The audios illustrated in the image are the English translations of the Mandarin Chinese used in the experiments.

After participants were comfortable with the experimental aim and the procedure, the experimenter then helped participants to perform the standard EyeLink calibration and validation routines.

Each trial involved two animals. The time line of a typical trial is summarized in Figure 3. Participants first saw two images of one animal each printed on the screen in turn, along with the



name of the animals played via the two loudspeakers situated at both sides of the screen. A black dot was then presented at the center of the screen. The participant was instructed to press the SPACE key while fixating on the dot. The press brought up the test image. The 500 ms after the onset of the test image, the test sentence began to play. The 4,000 ms after the offset of the test audio or pressing a key brought out a new trial. Participants' task was to determine which box the test sentence was talking about and pressed the corresponding key on the keyboard as soon as possible. Participants' eye movements were recorded from the onset of the test image to the offset of the trial.

## Results

### Behavioral Responses

The correct response to a conjunction was the big open box containing the tokens of both conjuncts. The correct response to a *but*-statement was the small open box containing the token of the first conjunct but not the second conjunct. Participants' behavioral responses to the disjunction, however, depended on whether the two inferences were processed. If participants computed neither the scalar implicature nor the ignorance inference, then all the four boxes in Figure 1 were eligible selections. If the scalar implicature but not the ignorance inference was computed, then the big-open box A would be ruled out, and the remaining three boxes B, C, and D were eligible options. If the ignorance inference but not the scalar implicature was computed, then the two small open boxes C and D would be ruled out, and boxes A and B would remain to be eligible options. And the small-closed box B would not be chosen unless

both the scalar implicature and the ignorance inference were computed.

The left panel of Figure 4 summarized participants' behavioral responses in Experiment 1. In our experimental design, the correct answers to the conjunctions were the big open-boxes (Box A), and the correct answers to the *but*-statements were the small-open boxes (Box D) containing the animals that were mentioned in the first conjuncts. Things became complex when the test sentences were disjunctions. In our experimental setting, all the boxes were compatible with the literal meaning of the uttered disjunctions, so participants' behavioral responses to the conjunctions could not be categorized as correct or incorrect. However, the boxes participants actually chose could inform us as to whether they computed the two proposed inferences or not. If participants computed the scalar implicature but not the ignorance inference, then they would choose boxes B, C or D. If participants computed the ignorance inference, then they would choose boxes A or B. If box B was the only choice participants made, it would suggest that participants computed both the scalar implicature and the ignorance inference. As we saw in the left panel of Figure 3, participants predominantly chose the big-open boxes, the first-mentioned small box, and the small-closed box, when they heard conjunctions, *but*-statements, and disjunctions, respectively. These findings suggested that both the scalar implicature and the ignorance inference were computed, and their computation was no later than the temporal location when participants overtly gave their behavioral responses.

However, the behavioral responses didn't tell us when the two inferences occurred while participants were listening to the

test audios. To explore how the test statements were processed online and when the two inferences occurred, we then analyzed participants' eye movements on the test images as they were listening to the auditorily presented test statements.

### Eye-Tracking Results

The test audios were 11 s long, and the eye-tracker had a sample rate of 500 Hz, so we had 5,500 sample points per testing trial. To process the eye-tracking data, we first deleted the samples where participants' eye movements were not caught, such as when they blinked their eyes. This process roughly affected 10% of the recorded data. We then defined four equal-sized areas of interest in the test image, containing the four boxes, respectively. Third, we then coded the recorded data as follows: for a specific area of interest, the samples where participants' fixations locating in that area was coded as 1, and the samples where participants' fixations locating out of that area was coded as 0.

The results of Experiment 1 were summarized in the left panel of **Figure 5**, where the *x*-axis was the sample point where the eye movement was recorded and *y*-axis was the proportion of samples where participants located their eye fixations on a specific area of interest. The four panels depicted participants' fixation patterns on the four areas of interest. The red, green, and blue lines illustrated participants' fixation patterns when the test statements were conjunctions, but-statements, and disjunctions, respectively. We were interested in how participants' fixation patterns were distributed among the four interest areas and how these distributions were influenced by hearing auditorily presented test sentences. The two dashed vertical lines illustrated the onset and offset of the sentential connectives. The horizontal dotted line labeled the proportion of 25%, illustrating the chance level of participants' fixation patterns. A preferred box should be fixated at more than 25% of the recorded samples.

To conduct the statistical analysis, we fitted a binomial generalized linear mixed model (GLMM) to the data at each interest area of each sample point. The GLMM model contained one fixed term: *sentential connectives*. The baseline of this fixed items differed when the analyzed interest area, i.e., the analyzed boxes, were different. To be specific, the conjunction was chosen as the baseline when analyzing the big-open box, the disjunction was chosen as the baseline when analyzing the small-closed box, and the *but*-statement was chosen as the baseline when analyzing the first-mentioned box (Box D). To summarize, the sentential connective whose expected response was the same as the analyzed interest area would be chosen as the baseline. The GLMM model included two random terms: *participants* and *items*. The formula evaluated to the two random terms included both the intercepts and the slope of the *sentential connectives*. The model fitting process was conducted via *glmer* function from *lme4* package (Bates et al., 2015) under the *R* environment (R Core Team, 2016). The *p*-values obtained using Wald *z*-tests were then Bonferroni adjusted. The gray areas in **Figure 4** signified the temporal periods where a significant difference existing between the baseline condition and the disjunctive condition ( $p < 0.05$ , Bonferroni adjusted). The statistical results with the wrong answered trials being excluded from the analyses were the same

as the statistical results with all trials being included, so we only reported the results with all trials being included in the analyses.

As we saw in the left panel of **Figure 5**, prior to the onset of the sentential connectives, no difference was observed between the three conditions, among all the four interest areas. These results provided reasonable bases that all the observed differences among the three conditions after the onset of the sentential connectives resulted from the sentential connectives, but not from other confounding factors. After the onset and prior to the offset of the sentential connectives, the disjunctive connective *or* had already triggered significantly different eye-movements than the baseline conditions among the three interest areas. And the starting points where a significant effect was first observed were summarized in **Table 2**. First, the disjunctive connective *or* had triggered significantly fewer fixations on the big-open box (Box A) than the conjunctive connective *and*; and this effect happened no later than 0.724 s after the onset of the sentential connectives. This suggested that the scalar implicature had already been fully processed prior to the offset of the sentential connectives. Second, the disjunctive connective *or* had also triggered significantly fewer fixations on the first-mentioned box (Box D) than the sentential connective *but*; and this effect happened no later than 0.714 s after the onset of the sentential connectives. This suggested that the ignorance inference had already been fully processed prior to the offset of the sentential connectives. Third, the disjunctive connective *or* had already triggered significantly more fixations on the small-closed box (Box B) than the sentential connectives *and* and *but*; and this effect happened no later than 0.690 s after the onset of the sentential connectives. Fourth, participants' fixations on the second-mentioned box (Box C) were never bigger than the chance level (0.25), suggesting that the second-mentioned box was never regarded as a legitimate option to our test audios. These results confirmed the previous two observations that both the scalar implicature and ignorance inference have already been fully processed prior to the offset of the sentential connectives.

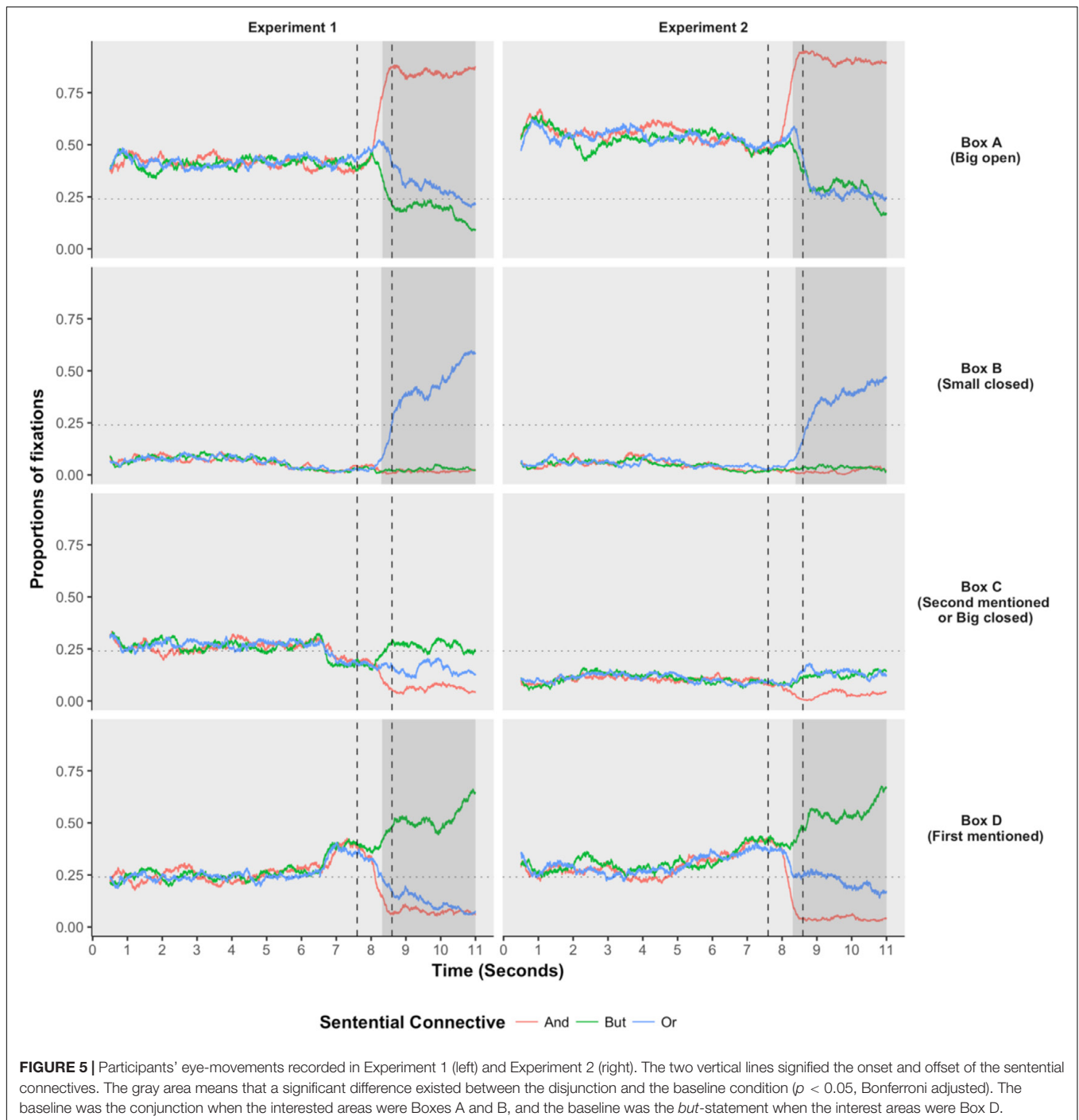
### Discussion

To summarize, Experiment 1 found that the scalar implicature and the ignorance inference were both locally computed. The effects triggered by the two inferences emerged no

**TABLE 2** | The temporal locations where significant differences started to be observed.

Experiment	Interest area	Baseline	Trial onset	Connective onset
One	Big open (Box A)	And	8.324	0.724
	Small closed (Box B)	And	8.290	0.690
	First mentioned (Box D)	But	8.314	0.714
Two	Big open (Box A)	And	8.304	0.704
	Small closed (Box B)	And	8.388	0.788
	First mentioned (Box D)	But	8.310	0.710

"Trial onset" illustrated the temporal locations based on the onset of the trial, i.e., the onset of the test image. "Connective onset" illustrated the temporal locations based on the onset of the sentential connectives.



later than the offset of the sentential connectives. The two inferences occurred almost at the same time when participants were listening to the test audios. These findings suggested that the radical grammatical account put forward by Meyer (2013) was more reasonable than the pragmatic account (Horn, 1972; Fauconnier, 1975; Sauerland, 2004; van Rooij and Schulz, 2004; Russell, 2006; Spector, 2007; Geurts, 2009; Franke, 2011) and hybrid account (Chierchia, 2004, 2017; Fox, 2007, 2014; Fox and Hackl, 2007; Magri, 2009,

2011; Chierchia et al., 2012) in terms of explaining our data.

However, the big open boxes in Experiment 1 were able to be ruled out both by the scalar implicature and by the ignorance inference that applied to the corresponding conjunctions. Experiment 1 could not be used to determine whether a weak inference (or ignorance inference) had been (temporarily or permanently) computed, because this experiment didn't contain a situation where the speaker was ignorant about the truth of the

corresponding conjunction. To solve this problem, Experiment 2 introduced a big-closed box, where the speaker and the hearer didn't know whether the corresponding conjunction was true or false. Under this experimental setting, if the computed inference to the corresponding conjunction is (temporarily or permanently) an ignorance inference, then the big open boxes, but not the big closed boxes would be ruled out. This would result in significant fewer fixations on the big open boxes, but not on the big closed boxes. In contrast, if the computed inference to the corresponding conjunction is a scalar implicature, then both the big open boxes and the big closed boxes would be ruled out. This would give rise to significantly fewer fixations on all big boxes, irrespective of whether they were open or closed. In Experiment 2, participants' fixations on the big closed boxes will be crucial to differentiate between the two possibilities.

## EXPERIMENT 2

### Method

#### Participants, Stimuli, and Procedure

Thirty-six postgraduate students from the Beijing Language and Culture University participated in this experiment. All the participants were native speakers of Mandarin Chinese, with normal or corrected to normal visions. None of these participants had participated in Experiment 1. They were paid 30CNY (approximately \$5) for their participation.

The stimuli and experimental procedure used in Experiment 2 were exactly the same as Experiment 1 with the following exception. In Experiment 1, a test image consisted of three small boxes and one big box. Two small boxes were open and contained two animals that were mentioned in the first and second merged propositions, respectively. In our example as illustrated in the left panel of **Figure 1**, the two-small open boxes were *C* and *D*, containing "cow" and "rooster," respectively. In Experiment 2, however, the small box containing the animal that was mentioned in the second merged proposition was replaced by a big closed box. In our example, the second mentioned animal "rooster" was contained in box *C*. In Experiment 2, the small open box *C* was replaced by a big closed one, yielding the right panel of **Figure 1**.

### Results and Discussion

Participants' behavioral responses in Experiment 2 (right panel of **Figure 4**) were similar to that observed in Experiment 1, indicating that replacing a small open box with a big closed box did not have a significant effect on participants' behavioral responses.

Participants' fixations patterns (right panel of **Figure 5**) and the onsets of the significant difference (**Table 2**) observed in Experiment 2 were also similar to that observed in Experiment 1. Furthermore, the big-closed box (Box *C*) always received fewer fixations than the chance level, regardless of the temporal positions and the sentential connectives, suggesting that a big box was never regarded as a valid option of the

disjunction, regardless of whether the big box was open or closed.

## GENERAL DISCUSSION

First, our results have important implications for adjudicating between different accounts of scalar implicatures and ignorance inferences. The pragmatic account (Horn, 1972; Fauconnier, 1975; Levinson, 2000; Sauerland, 2004; Russell, 2006; Schulz and van Rooij, 2006; Spector, 2007; Geurts, 2009, 2010; Franke, 2011) regards both inferences as pragmatic processes. A pragmatic process applies over speech acts while a speech act is derived from the whole statement, and as pragmatic processes, the two inferences are not expected to occur until the offsets of the test sentences (but see, Chemla and Singh, 2014a,b, for different viewpoints). Furthermore, the pragmatic account predicts that the ignorance inference should occur earlier than the scalar implicature (Chemla and Singh, 2014a,b). Our results are contradictory to their predictions. We observed that both the scalar implicature and the ignorance inference were computed prior to the offset of the sentential connectives *or* that triggered the two inferences. These two inferences occurred almost at the same time. These results are also contradictory to the hybrid account (Chierchia, 2004, 2017; Fox, 2007, 2014; Fox and Hackl, 2007; Magri, 2009, 2011; Chierchia et al., 2012), which regards the scalar implicature as a grammatical process, but regards the ignorance inference as a pragmatic process. A grammatical process should occur earlier than that of a pragmatic process. According to the hybrid theory, the scalar implicature should be locally computed, but the ignorance inference should not. The scalar implicature is expected to be processed earlier than that of the ignorance inference. Our results are in a par with the radically grammatical account (Meyer, 2013), which regards both the scalar implicature and the ignorance inference as grammatical processes. The two grammatical inferences are triggered by the same lexical item, i.e., the disjunctive connective *or* in our experiments, so the two inferences are expected to occur at the same time.

Second, Experiment 2 explored whether or not there is an intermediate stage between the literal meaning and the scalar implicature. This stage is called the primary inference by the pragmatic account. The findings showed that there is no such intermediate stage. These findings are contradictory to the pragmatic account (Horn, 1972; Fauconnier, 1975; Levinson, 2000; Sauerland, 2004; van Rooij and Schulz, 2004; Russell, 2006; Schulz and van Rooij, 2006; Spector, 2007; Geurts, 2009, 2010; Franke, 2011), but are consistent with both the hybrid account (Chierchia, 2004, 2017; Fox, 2007, 2014; Fox and Hackl, 2007; Magri, 2009, 2011; Chierchia et al., 2012) and the second reading of the radically grammatical account (Meyer, 2013) that the scalar implicature is not derived the maxims of conversations.

Third, there exist several upper-bounded construals that engender scalar implicatures, including the disjunctive connective *or* explored here and the existential quantifier

*some* explored extensively in literature. The scales engendered by different scalar expressions are traditionally regarded as having the same properties. Recently, researchers have begun to realize that there might be a heterogeneity between different scalar scales (Doran et al., 2009, 2012; van Tiel et al., 2016). For example, the distinctness of the scalemates have been found to affect participants behavioral responses (van Tiel et al., 2016). Our experiments found that the disjunctive connective *or* induces both a scalar implicature to its Horn scales and an ignorance inference to its two disjuncts, which is different from other scalar expressions. Any robust theory of quantity-based implicatures should encompass the variety between different scalar expressions. Regardless of the varieties, these scalar expressions are not necessarily different from each other on the pragmatic-grammatical dimension. And the distinction on this dimension is crucial in differentiating different accounts. Previous studies (Breheny et al., 2006, 2013; Grodner et al., 2010; Degen and Tanenhaus, 2015; Foppolo and Marelli, 2017) suggest that the scalar implicature computed from the quantifier *some* is a semantic process. The two experiments I reported here also supported the idea that both the scalar implicature and the ignorance inference engendered from the disjunctive connective *or* are semantic processes. Furthermore, our preliminary data exploring the online processing of the modal verbs [*might, must*] and of the quantificational adverbs [*sometimes, always*] suggest that these scales also are immediately constructed once the modal verbs and the quantificational adverbs are encountered. Taken together, the available online processing

data using the visual world paradigm tend to support the idea that the scalar implicature in general is a grammatical process.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Beijing Language and Culture University Committee on Human Research Protection with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Beijing Language and Culture University Committee on Human Research Protection.

## AUTHOR CONTRIBUTIONS

LZ designed, conducted the experiments, analyzed the data, and wrote the paper.

## ACKNOWLEDGMENTS

This research was supported by Two Fundamental Research Funds for the Central Universities [grant numbers: 15YBB29 and 15YJ050003].

## REFERENCES

- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bergen, L., and Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 1450–1460. doi: 10.1037/a0027850
- Bott, L., and Noveck, I. A. (2004). Some utterances are underinformative: the onset and time course of scalar inferences. *J. Mem. Lang.* 51, 437–457. doi: 10.1016/j.jml.2004.05.006
- Breheny, R., Ferguson, H. J., and Katsos, N. (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Lang. Cogn. Process.* 28, 443–467. doi: 10.1080/01690965.2011.649040
- Breheny, R., Katsos, N., and Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition* 100, 434–463. doi: 10.1016/j.cognition.2005.07.003
- Chemla, E., Cummins, C., and Singh, R. (2017). Training and timing local scalar enrichments under global pragmatic pressures. *J. Semant.* 34, 107–126. doi: 10.1093/jos/ffw006
- Chemla, E., and Singh, R. (2014a). Remarks on the experimental turn in the study of scalar implicature, part I. *Lang. Linguist. Compass* 8, 373–386. doi: 10.1111/lnc3.12081
- Chemla, E., and Singh, R. (2014b). Remarks on the experimental turn in the study of scalar implicature, part II. *Lang. Linguist. Compass* 8, 387–399. doi: 10.1111/lnc3.12080
- Chemla, E., and Spector, B. (2011). Experimental evidence for embedded scalar implicatures. *J. Semant.* 28, 359–400. doi: 10.1093/jos/ffq023
- Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V., and Sperber, D. (2008). Making disjunctions exclusive. *Q. J. Exp. Psychol.* 61, 1741–1760. doi: 10.1080/17470210701712960
- Chierchia, G. (2004). “Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface,” in *Structures and Beyond: The Cartography of Syntactic Structures*, Vol. 3, ed. A. Belletti (New York, NY: Oxford University Press), 39–103.
- Chierchia, G. (2017). Scalar implicatures and their interface with grammar. *Annu. Rev. Linguist.* 3, 245–264. doi: 10.1146/annurev-linguistics-011516-033846
- Chierchia, G., Danny Fox, C., and Spector, B. (2012). “Scalar implicature as a grammatical phenomenon,” in *Semantics: An International Handbook of Natural Language Meaning*, eds C. Maienborn, K. V. Heusinger, and P. Portner (Berlin: De Gruyter Mouton), 2297–2331.
- Clifton, C. Jr., and Dube, C. (2010). Embedded implicatures observed: a comment on Geurts and Poussoulous (2009). *Semant. Pragmat.* 3, 1–13. doi: 10.3765/sp.3.7
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cogn. Psychol.* 6, 84–107. doi: 10.1016/0010-0285(74)90005-x
- Degen, J., and Tanenhaus, M. K. (2015). Processing scalar implicature: a constraint-based approach. *Cogn. Sci.* 39, 667–710. doi: 10.1111/cogs.12171
- Doran, R., Baker, R., McNabb, Y., Larson, M., and Ward, G. (2009). On the non-unified nature of scalar implicature: an empirical investigation. *Int. Rev. Pragmat.* 1, 211–248. doi: 10.1163/187730909X12538045489854
- Doran, R., Ward, G., Larson, M., McNabb, Y., and Baker, R. E. (2012). A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language* 88, 124–154. doi: 10.1353/lan.2012.0008
- Fauconnier, G. (1975). Pragmatic scales and logical structure. *Linguist. Inq.* 6, 353–375.
- Foppolo, F., and Marelli, M. (2017). No delay for some inferences. *J. Semant.* 34, 659–681. doi: 10.1093/jos/ffx013

- Fox, D. (2007). "Free choice and the theory of scalar implicatures," in *Presupposition and Implicature in Compositional Semantics*, eds U. Sauerland and P. Stateva (New York, NY: Palgrave Macmillan), 71–120. doi: 10.1057/9780230210752\_4
- Fox, D. (2014). Cancelling the maxim of quantity: another challenge for a Gricean theory of scalar implicatures. *Semant. Pragmat.* 7, 1–20. doi: 10.3765/sp.7.5
- Fox, D., and Hackl, M. (2007). The universal density of measurement. *Linguist. Philos.* 29, 537–586. doi: 10.1007/s10988-006-9004-4
- Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semant. Pragmat.* 4, 1–82. doi: 10.3765/sp.4.1
- Geurts, B. (2009). Scalar implicature and local pragmatics. *Mind Lang.* 24, 51–79. doi: 10.1111/j.1468-0017.2008.01353.x
- Geurts, B. (2010). *Quantity Implicature*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511975158
- Geurts, B., and Pouscoulous, N. (2009). Embedded implicatures?? *Semant. Pragmat.* 2, 1–34. doi: 10.3765/sp.2.4
- Grice, H. P. (1975). "Logic and conversation," in *Syntax and Semantics*, Vol. 3, eds P. Cole and J. L. Morgan (New York, NY: Academic Press), 41–58.
- Grodner, D. J., Klein, N. M., Carbary, K. M., and Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: evidence for immediate pragmatic enrichment. *Cognition* 116, 42–55. doi: 10.1016/j.cognition.2010.03.014
- Horn, L. R. (1972). *On the Semantic Properties of Logical Operators in English*. Ph.D. thesis, University of California, Los Angeles, CA.
- Huang, Y. T., and Snedeker, J. (2009). Online interpretation of scalar quantifiers: insight into the semantics–pragmatics interface. *Cogn. Psychol.* 58, 376–415. doi: 10.1016/j.cogpsych.2008.09.001
- Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature, Language, Speech, and Communication*. Cambridge, MA: The MIT Press.
- Magri, G. (2009). A theory of individual-level predicates based on blind mandatory scalar implicatures. *Nat. Lang. Semant.* 17, 245–297. doi: 10.1007/s11050-009-9042-x
- Magri, G. (2011). Another argument for embedded scalar implicatures based on oddness in downward entailing environments. *Semant. Pragmat.* 4, 1–51. doi: 10.3765/sp.4.6
- Marty, P. P., and Chemla, E. (2013). Scalar implicatures: working memory and a comparison with only. *Front. Psychol.* 4:403. doi: 10.3389/fpsyg.2013.00403
- Meyer, M.-C. (2013). *Ignorance and Grammar*. Ph.D. thesis, Massachusetts Institute Of Technology, Cambridge, MA.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Russell, B. (2006). Against grammatical computation of scalar implicatures. *J. Semant.* 23, 361–382. doi: 10.1093/jos/ffl008
- Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguist. Philos.* 27, 367–391. doi: 10.1023/B:LING.0000023378.71748.db
- Sauerland, U. (2012). The computation of scalar implicatures: pragmatic, lexical or grammatical? *Lang. Linguist. Compass* 6, 36–49. doi: 10.1002/lnc3.321
- Schulz, K., and van Rooij, R. (2006). Pragmatic meaning and non-monotonic reasoning: the case of exhaustive interpretation. *Linguist. Philos.* 29, 205–250. doi: 10.1007/s10988-005-3760-4
- Spector, B. (2007). "Scalar implicatures: exhaustivity and gricean reasoning?," in *Questions in Dynamic Semantics*, eds M. Aloni, A. Butler, and P. Dekker (Oxford: Elsevier), 225–249.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634. doi: 10.1126/science.7777863
- van Rooij, R., and Schulz, K. (2004). Exhaustive interpretation of complex sentences. *J. Logic Lang. Inform.* 13, 491–519. doi: 10.1007/s10849-004-2118-6
- van Tiel, B., van Miltenburg, E., Zevakhina, N., and Geurts, B. (2016). Scalar diversity. *J. Semant.* 33, 137–175. doi: 10.1093/jos/ffu017
- Zhan, L., Crain, S., and Zhou, P. (2015). The online processing of only if- and even if- conditional statements: implications for mental models. *J. Cogn. Psychol.* 26, 367–379. doi: 10.1080/20445911.2015.1016527

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Zhan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# ULTRA: Universal Grammar as a Universal Parser

David P. Medeiros\*

Department of Linguistics, University of Arizona, Tucson, AZ, United States

## OPEN ACCESS

### Edited by:

Ángel J. Gallego,  
Universitat Autònoma de Barcelona,  
Spain

### Reviewed by:

Ricardo Etxepare,  
UMR5478 Centre de Recherche sur la  
Langue et les Textes Basques (IKER),  
France

Cristiano Chesì,  
Istituto Universitario di Studi Superiori  
di Pavia (IUSS), Italy

### \*Correspondence:

David P. Medeiros  
medeiros@email.arizona.edu

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 31 August 2017

**Accepted:** 30 January 2018

**Published:** 15 February 2018

### Citation:

Medeiros DP (2018) ULTRA: Universal  
Grammar as a Universal Parser.  
*Front. Psychol.* 9:155.  
doi: 10.3389/fpsyg.2018.00155

A central concern of generative grammar is the relationship between hierarchy and word order, traditionally understood as two dimensions of a single syntactic representation. A related concern is directionality in the grammar. Traditional approaches posit process-neutral grammars, embodying knowledge of language, put to use with infinite facility both for production and comprehension. This has crystallized in the view of Merge as the central property of syntax, perhaps its only novel feature. A growing number of approaches explore grammars with different directionalities, often with more direct connections to performance mechanisms. This paper describes a novel model of universal grammar as a one-directional, universal parser. Mismatch between word order and interpretation order is pervasive in comprehension; in the present model, word order is language-particular and interpretation order (i.e., hierarchy) is universal. These orders are not two dimensions of a unified abstract object (e.g., precedence and dominance in a single tree); rather, both are temporal sequences, and UG is an invariant real-time procedure (based on Knuth's stack-sorting algorithm) transforming word order into hierarchical order. This shift in perspective has several desirable consequences. It collapses linearization, displacement, and composition into a single performance process. The architecture provides a novel source of brackets (labeled unambiguously and without search), which are understood not as part-whole constituency relations, but as storage and retrieval routines in parsing. It also explains why neutral word order within single syntactic cycles avoids 213-like permutations. The model identifies cycles as extended projections of lexical heads, grounding the notion of phase. This is achieved with a universal processor, dispensing with parameters. The empirical focus is word order in noun phrases. This domain provides some of the clearest evidence for 213-avoidance as a cross-linguistic word order generalization. Importantly, recursive phrase structure "bottoms out" in noun phrases, which are typically a single cycle (though further cycles may be embedded, e.g., relative clauses). By contrast, a simple transitive clause plausibly involves two cycles (vP and CP), embedding further nominal cycles. In the present theory, recursion is fundamentally distinct from structure-building within a single cycle, and different word order restrictions might emerge in larger domains like clauses.

**Keywords:** syntax, linearization, parsing, universal grammar, word order, typology, universal 20, stack-sorting

## INTRODUCTION

One of the most significant recent developments for linguistic theory is the appearance of high-quality datasets on the full range of cross-linguistic variation. In the past, generative studies typically relied on detailed examination of one or several languages to illuminate syntactic mechanisms. While this approach is certainly fruitful, the accumulation of information about large numbers of languages opens new possibilities for sharpening understanding.

Within generative grammar, considerable attention has been given to recursion as a (or even *the*) fundamental property of language (see Berwick and Chomsky, 2016 for discussion). This is formalized in a core operation called Merge, combining two syntactic objects (ultimately built from lexical items) into a set containing both. Recursion follows from the ability of Merge to apply to its own output. Merge also captures the essential fact that sentences have internal structure (bracketed constituency), each layer corresponding to an application of Merge.

Contrary to this framework, I argue that it is a conceptual error to view sentences as groupings (whether sets, or something else) of lexical items. The error inheres in thinking of lexical items as coherent units existing at a single level. This leads to thinking of sentences as single-level representations as well. Words, put simply, aren't *things*; they are a pair of processes, extended in time. In the context of comprehension, the relevant processes are recognition of the word, and integration of its meaning into an interpretation. I develop a novel view of the structure of sentences in terms of these two kinds of processes. Crucially, a non-trivial relationship governs their relative sequencing: one word may occur earlier than another in surface order, yet its meaning may be integrated later. Considering sentences as unified, atemporal representations built atop impenetrable lexical atoms leaves us unable to capture the fundamentally temporal phenomena involved, in which the two aspects of each word are not bundled together, and the processes for different words interweave.

This paper proposes a novel model of grammatical mechanisms, called ULTRA (Universal Linear Transduction Reactive Automaton). Within local syntactic domains forming the extended projection of a lexical root (such as a verb or noun), ULTRA employs Knuth's (1968) stack-sorting algorithm to directly map surface word orders to underlying base structure. The mapping succeeds only for 213-avoiding orders. This is an intriguing result, as 213-avoidance arguably bounds neutral word order variation across languages, in a variety of syntactic domains. While the local sound and meaning representations in this model are sequences, hierarchical structure nevertheless arises in the dynamic action of the mapping. The bracketed structures found here, although epiphenomenal, closely match those built by Merge, with some crucial differences (arguably favoring the present theory).

Stack-sorting proves to be an effective procedure for linking word order and hierarchical interpretation, encompassing linearization, displacement, composition, and labeled brackets. The theory invites realization as a real-time performance process. Pursuing that realization significantly recasts the

boundaries between performance and competence. Remarkably, ULTRA requires no language-particular parameters; an invariant algorithm serves as grammatical device for all languages. Put simply, I propose that Universal Grammar is a universal parser.

Nevertheless, stack-sorting is too limited a mechanism to describe all the phenomena of human syntax. Three kinds of effects are left hanging: unbounded recursion, non-neutral orders, and the existence of apparently distinct languages. Moreover, understanding stack-sorting as a processing system encounters two obvious problems: it is a unidirectional parser, not trivially reversible for production; and it conflicts with strong evidence for word-by-word incrementality in comprehension.

Although constructing a complete model of syntax and processing goes far beyond the scope of the paper, the problems that arise in basing a parser-as-grammar model on stack-sorting warrant consideration. I appeal to the distinction between reactive and predictive processes, casting stack-sorting as a universal reactive routine. A separate predictive module plays a crucial role in production, and in the appearance of distinct, relatively rigid word orders. Prediction also helps reconcile ULTRA with incremental interpretation. I appeal to properties of memory to resolve further problems, speculating that primacy memory (distinct from the recency memory underpinning stack-sorting) is the source of another cluster of syntactic properties, including long-distance movement, crossing dependencies, and the special syntax of the "left periphery." Finally, I suggest that episodic memory—independently hierarchical in structure, in humans—plays a key role in linguistic recursion.

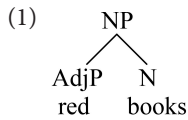
The structure of this paper is as follows. Section Linear Base argues that the "base" structure within each local syntactic domain is a sequence. Section \*213 in Neutral Word Order explores the generalization that 213-avoidance delimits information-neutral word order possibilities, across languages. Section Stack-Sorting as a Grammatical Mechanism proposes a stack-sorting procedure to capture 213-avoidance in word order. Section Stack-Sorting: Linearization, Displacement, Composition, and Labeled Brackets shows how further syntactic effects follow from stack-sorting. Section Comparison with Existing Accounts of Universal 20 compares ULTRA to existing accounts of 213-avoidance in word order, focusing on Universal 20. Section Universal Grammar as Universal Parser pursues the realization of stack-sorting in real-time performance. Section Possible Extensions to a More Complete Theory of Syntax addresses the challenges in taking stack-sorting as the core of Universal Grammar, sketching some possible extensions. Section Conclusion concludes.

## LINEAR BASE

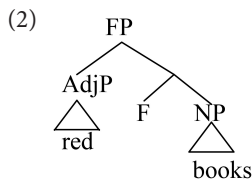
Syntactic combination could take many forms. An emerging view is that combination largely keeps to head-complement relations (Starke, 2004; Jayaseelan, 2008). The term "head" has at least two different senses, in this context. First, in any combination of two syntactic objects, one is "more central" to the composite meaning. Let us call this notion of head the *root*, noting that in extended projections of nouns and verbs, the lexical noun or verb root is

semantically dominant. The other sense of head concerns which element determines the combinatoric behavior of the composite; let us call this notion of head the *label*.

In older theories of phrase structure, the two senses of head (root and label) converged on the same element; a noun, for example, combined with all its modifiers within a noun phrase. Headedness thus mapped to hierarchical dominance; the root projected its label above its dependents. To illustrate, a combination of adjective and noun, such as *red books*, would be represented as follows.



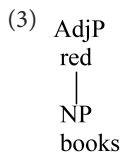
This traditional conclusion about the relationship of dependency and hierarchy is overturned in modern syntactic cartography (Rizzi, 1997; Cinque, 1999, and subsequent work). Cartographic approaches propose that syntactic combination follows a strict, cross-linguistically uniform hierarchy, within each extended projection. This hierarchy involves a sequence of functional heads, licensing combination with various modifiers in rigid order. The phrase *red books* is represented as follows.



Here, the adjective is the specifier of a dedicated functional head (F), which labels the composite, determining its combinatoric behavior. In cartographic representations heads are uniformly *below* their dependents, which appear higher up the spine.

Questions arise about these representations, which postulate an abundance of unpronounced material. A curious observation is that functional heads and their specifiers seem not to occur together overtly, as formalized in Koopman’s (2000) Generalized Doubly-Filled Comp Filter.

Starke (2004) takes Koopman’s observation further, arguing that heads and specifiers do not co-occur because they are *tokens of the same type*, competing for a single position. Starke recasts the cartographic spine as an abstract functional sequence (*fseq*), whose positions can be discharged equally by lexical or phrasal material. Pursuing Starke’s conception, the adjective-noun combination would be represented as below.



Again, we have reversed traditional conclusions about the hierarchy of heads and dependents. Nevertheless, the notion of root (picking out the noun) is still crucial, as the modifiers occur in the hierarchical order dictated by *its* fseq.

Syntactic combination of this sort is sequential, within each extended projection. These “base” sequences encode bottom-up composition, so it is natural to order the sequence in the same way (bottom-up). The base (i.e., fseq, cartographic spine) is widely taken to be uniform across languages, and to express “thematic,” information-neutral meaning (contrasted with discourse-information structure)<sup>1</sup>.

A grammar, on anyone’s theory, specifies a formal mapping linking sound and meaning (more accurately, outer and inner form, allowing for non-auditory modalities). This specification could take many forms. Sequential representation of the base allows a remarkably simple formulation of the sound-meaning mapping. This reformulation yields a principled account of a class of word order universals. Moreover, while the interface objects (word orders, and base trees) involved in the mapping are sequences, bracketed hierarchical structure arises as a dynamic effect.

There are various ways of conceptualizing the relationship between the base and surface word order. The usual view is that the base orders the input to a derivation, yielding surface word order as the output. That directionality is implicit in terms used to describe the hierarchy-order relation: *linearization*, *externalization*, etc. This paper pursues a different view, where surface word orders are inputs to an algorithm that attempts to assemble the base as output. Significantly, the only inputs that converge on the uniform base under this process are 213-avoiding; all 213-containing word orders result in deviant output.

### \*213 IN NEUTRAL WORD ORDER

213-avoidance arguably captures information-neutral word order possibilities in a variety of syntactic domains, across languages. By 213-avoidance, I mean a ban on surface order ...*b*...*a*...*c*..., for elements  $a \gg b \gg c$ , where  $\gg$  indicates c-command in standard tree representations of the base (equivalently, dominance in Starke’s trees). In other words, neutral word orders seem to avoid a mid-high-low (sub)sequence of elements from a single fseq. The elements forming this forbidden contour need not be adjacent, in surface order or in the base fseq.

213-avoidance is widely believed to delimit the ordering options for verb clusters, well-known in West Germanic (see Wurmbrand, 2006 for an overview). Barbiers et al.’s (2008) extensive survey of Dutch dialects found very few instances of this order; German dialects seem to avoid this order as well<sup>2</sup>.

<sup>1</sup>This underlies Chomsky’s claim that the distinction between External Merge (the base) and Internal Merge (displacement) correlates with the Duality of Semantics: “External Merge correlates with argument structure, internal Merge with edge properties, scopal or discourse-related (new and old information, topic, etc.)” (Chomsky, 2005, p. 14). However, some neutral word orders require Internal Merge to derive (even allowing free linearization of sister nodes; see Abels and Neeleman, 2012). ULTRA maintains the identification of the base with thematic structure, while rejecting the empirically problematic claim that displacement gives rise to scopal and discourse-information properties.

<sup>2</sup>Schmid and Vogel (2004) report examples of this order in German dialects, but note that focus seems to be involved. Intriguingly, many instances of 213 order are only felicitous under special discourse-information conditions. However,

Meanwhile Zwart (2007) analyzes 213 order in Dutch verb clusters as involving extraposition of the final element<sup>3</sup>.

The best-studied domain supporting 213-avoidance in word order is Greenberg’s Universal 20, describing noun phrase orders.

“When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in that order. If they follow, the order is either the same or its exact opposite” (Greenberg, 1963 p. 87).

Subsequent work has refined this picture. Cinque (2005) reports that only 14 of 24 logically possible orders of these elements are attested as information-neutral orders (Table 1).

Cinque describes these facts with a constraint on movement from a uniform base. Specifically, he proposes that all movements move the noun, or something containing it, to the left (section Comparison with Existing Accounts of Universal 20 details Cinque’s theory and related accounts). What is forbidden is remnant movement.

Noun phrase orders obey a simple generalization: attested orders are all and only 213-avoiding permutations. All unattested orders have 213-like subsequences. For example, unattested \*Num Dem Adj N contains subsequences Num Dem Adj..., and Num Dem ... N, representing mid-high-low contours with respect to the fseq.

## STACK-SORTING AS A GRAMMATICAL MECHANISM

There is a particularly simple procedure that maps 213-avoiding word<sup>4</sup> orders to the uniform base, called stack-sorting<sup>5</sup>. I describe an adaptation of Knuth (1968) stack-sorting algorithm, which

uses last-in, first out (stack) memory to sort items by their relative order in the base. This is a partial sorting algorithm: it only achieves the desired output for some input orders.

<p>(4) <b>STACK-SORTING ALGORITHM</b></p> <p>While input is non-empty,              If I ≫ S, Pop.              Else Push.</p> <p>While Stack is non-empty,              Pop.</p>	<p><b>DEFINITIONS</b></p> <p>I: next item in input.          S: item on top of stack.          x ≫ y: x c-commands y in the base (e.g., Dem ≫ N).          Push: moves I from input onto stack.          Pop: moves S from stack to output.</p>
---	---

(4) maps all and only 213-avoiding word orders to a 321-like hierarchy, corresponding to the base. 213-containing orders are mapped to a deviant output, distinct from the base. By hypothesis, that is why such orders are typologically unavailable: they are automatically mapped to an uninterpretable order of composition. This explains the Universal 20 pattern (Tables 2, 3).

Let us illustrate how (4) parses some noun phrase orders: Dem-Adj-N, N-Dem-Adj, \*Adj-Dem-Num.

- (5) Dem-Adj-N: PUSH(Dem), PUSH(Adj), PUSH(N), POP(N), POP(Adj), POP(Dem).
- (6) N-Dem-Adj: PUSH(N), POP(N), PUSH(Dem), PUSH(Adj), POP(Adj), POP(Dem).

For attested orders, the nominal categories POP in the order <N, Adj, Dem>, matching their bottom-up hierarchy.

- (7) \*Adj-Dem-N: PUSH(Adj), POP(Adj), PUSH(Dem), PUSH(N), POP(N), POP(Dem).

For the unattested 213-like order, items POP in the deviant order \* <Adj, N, Dem>, failing to construct the universal interpretation order.

That’s nice: (4) maps attested orders to their universal meaning, simultaneously ruling out unattested orders. But beyond such a mapping, an adequate grammar must explain other aspects of knowledge of language, including surface structure bracketing. If grammar treats surface orders and base structures as sequences<sup>6</sup> (locally), where can such bracketed structure come from?

**TABLE 1** | Possible noun phrase orders. Cinque (2005, pp. 319–320) report of the number of languages exhibiting each order is given by a number: 0 = unattested; 1 = very few languages; 2 = few languages; 3 = many languages; 4 = very many languages.

Dem Num Adj N 4	Dem Num N Adj 3	Dem N Num Adj 1	N Dem Num Adj 2
*Num Dem Adj N 0	*Num Dem N Adj 0	*Num N Dem Adj 0	*N Num Dem Adj 0
*Adj Dem Num N 0	*Adj Dem N Num 0	Adj N Dem Num 1	N Adj Dem Num 2
*Dem Adj Num N 0	Dem Adj N Num 1	Dem N Adj Num 3	N Dem Adj Num 1
*Num Adj Dem N 0	Num Adj N Dem 1	Num N Adj Dem 2	N Num Adj Dem 2
*Adj Num Dem N 0	*Adj Num N Dem 0	Adj N Num Dem 1	N Adj Num Dem 4

Cells with unattested orders are shaded for additional clarity. Attested orders are all and only the 213-avoiding permutations of the Dem ≫ Adj ≫ Num ≫ N base.

Salzmann has recently described neutral 213 verb cluster orders in Swiss German. I leave this possible counter-example to future investigation.

<sup>3</sup>Verb clusters are an instance of Restructuring, whereby multiple clauses are treated syntactically as monoclausal. Extraposition places the extraposed element in a separate domain. Zwart’s observation thus allows us to maintain the generalization that single-domain neutral orders are 213-avoiding.

<sup>4</sup>By focusing on word order, I am also setting aside morphological ordering and features. While I cannot pursue the issue here, there is evidence that morphology obeys similar cross-linguistic restrictions, and there is no reason why the sorting procedure could not apply to sub-word units.

<sup>5</sup>Stack-sorting is usually described as 231-avoiding. However, linguists effectively number their hierarchies backwards, assigning the highest number to the bottom of the hierarchy, the first element interpreted.

**TABLE 2** | Result of stack-sorting logically possible orders of 4 elements, in the format **input** → output.

1234 → 4321	1243 → 4321	1423 → 4321	4123 → 4321
*2134 → *2431	*2143 → *2431	*2413 → *4231	*4213 → *4231
*3124 → *3421	*3142 → *3421	3412 → 4321	4312 → 4321
*1324 → *3421	1342 → 4321	1432 → 4321	4132 → 4321
*2314 → *3241	2341 → 4321	2431 → 4321	4231 → 4321
*3214 → *3241	*3241 → *3421	3421 → 4321	4321 → 4321

213-avoiding orders (white cells) are stack-sorted into the 4,321 base sequence. Note that the correctly stack-sorted orders correspond exactly to the attested noun phrase orders, as reported by Cinque (2005).

<sup>6</sup>In formal language theory terms, stack-sorting is a kind of linear transduction. Linear transduction has largely been ignored as a possible model of grammar, in part because it seemed incapable of describing the hierarchical structure of linguistic expressions. Some researchers (e.g., Marco Kuhlmann and Markus Saers)

**TABLE 3 |** Stack sorting computations for 4-orders.

Output ← Stack ← Input			Output ← Stack ← Input			Output ← Stack ← Input			Output ← Stack ← Input						
Start	-	-	1234	Start	-	-	1243	Start	-	-	1423	Start	-	-	4123
Push 1		1	234	Push 1		1	243	Push 1		1	423	Push 4		4	123
Push 2		2/1	34	Push 2		2/1	43	Push 4		4/1	23	Pop 4	4		123
Push 3		3/2/1	4	Push 4		4/2/1	3	Pop 4	4	2/1	23	Push 1	4	1	23
Push 4		4/3/2/1		Pop 4	4	2/1	3	Push 2	4	2/1	3	Push 2	4	2/1	3
Pop 4	4	3/2/1		Push 3	4	3/2/1		Push 3	4	3/2/1		Push 3	4	3/2/1	
Pop 3	43	2/1		Pop 3	43	2/1		Pop 3	43	2/1		Pop 3	43	2/1	
Pop 2	432	1		Pop 2	432	1		Pop 2	432	1		Pop 2	432	1	
Pop 1	<b>4321</b>			Pop 1	<b>4321</b>			Pop 1	<b>4321</b>			Pop 1	<b>4321</b>		
Output ← Stack ← Input			Output ← Stack ← Input			Output ← Stack ← Input			Output ← Stack ← Input						
Start	-	-	*2134	Start	-	-	*2143	Start	-	-	*2413	Start	-	-	*4213
Push 2		2	134	Push 2		2	143	Push 2		2	413	Push 4		4	213
Pop 2	2		134	Pop 2	2		143	Push 4		4/2	13	Pop 4	4		213
Push 1	2	1	34	Push 1	2	1	43	Pop 4	4	2	13	Push 2	4	2	13
Push 3	2	3/1	4	Push 4	2	4/1	3	Pop 2	42		13	Pop 2	42		13
Push 4	2	4/3/1		Pop 4	24	1	3	Push 1	42	1	3	Push 1	42	1	3
Pop 4	24	3/1		Push 3	24	3/1		Push 3	42	3/1		Push 3	42	3/1	
Pop 3	243	1		Pop 3	243	1		Pop 3	423	1		Pop 3	423	1	
Pop 1	<b>*2431</b>			Pop 1	<b>*2431</b>			Pop 1	<b>*4231</b>			Pop 1	<b>*4231</b>		
Output ← Stack ← Input			Output ← Stack ← Input			Output ← Stack ← Input			Output ← Stack ← Input						
Start	-	-	*3124	Start	-	-	*3142	Start	-	-	3412	Start	-	-	4312
Push 3		3	124	Push 3		3	142	Push 3		3	412	Push 4		4	312
Pop 3	3		124	Pop 3	3		142	Push 4		4/3	12	Pop 4	4		312
Push 1	3	1	24	Push 1	3	1	42	Pop 4	4	3	12	Push 3	4	3	12
Push 2	3	2/1	4	Push 4	3	4/1	2	Pop 3	43		12	Pop 3	43		12
Push 4	3	4/2/1		Pop 4	34	1	2	Push 1	43	1	2	Push 1	43		1
Pop 4	34	2/1		Push 2	34	2/1		Push 2	43	2/1		Push 2	43	2/1	
Pop 2	342	1		Pop 2	342	1		Pop2	432	1		Pop 2	432	1	
Pop 1	<b>*3421</b>			Pop 1	<b>*3421</b>			Pop 1	<b>4321</b>			Pop 1	<b>4321</b>		
Output ← Stack ← Input			Output ← Stack ← Input			Output ← Stack ← Input			Output ← Stack ← Input						
Start	-	-	*1324	Start	-	-	1342	Start	-	-	1432	Start	-	-	4231
Push 1		1	324	Push 1		1	342	Push 1		1	431	Push 4		4	132
Push 3		3/1	24	Push 3		3/1	42	Push 4		4/1	32	Pop 4	4		132
Pop 3	3	1	24	Push 4		4/3/1	2	Pop 4	4	1	32	Push 1	4	1	32
Push 2	3	2/1	4	Pop 4	4	3/1	2	Push 3	4	3/1	2	Push 3	4	3/1	2
Push 4	3	4/2/1		Pop 3	43	1	2	Pop 3	43	1	2	Pop 3	43	2/1	2
Pop 4	34	2/1		Push 2	43	2/1		Push 2	432	2/1		Push 2	43	1	
Pop 2	342	1		Pop 2	432	1		Pop 2	432	1		Pop 2	432		
Pop 1	<b>*3421</b>			Pop 1	<b>4321</b>			Pop 1	<b>4321</b>			Pop 1	<b>4321</b>		
Output ← Stack ← Input			Output ← Stack ← Input			Output ← Stack ← Input			Output ← Stack ← Input						
Start	-	-	*2314	Start	-	-	2341	Start	-	-	2431	Start	-	-	4321
Push 2		2	314	Push 2		2	341	Push 2		2	431	Push 4		4	231
Push 3		3/2	14	Push 3		3/2	41	Push 4		4/2	31	Pop 4	4		231
Pop 3	3	2	14	Push 4		4/3/2	1	Pop 4	4	2	31	Push 2	4	2	31
Pop 2	32		14	Pop 4	4	3/2	1	Push 3	4	3/2	1	Push 3	4	3/2	1
Push 1	32	1	4	Pop 3	43	2	1	Pop 3	43	2	1	Pop 3	43	2	1
Push 4	32	4/1		Pop 2	432		1	Pop 2	432		1	Pop 2	432		1
Pop 4	324	1		Push 1	432	1		Push 1	432	1		Push 1	432	1	
Pop 1	<b>*3241</b>			Pop 1	<b>4321</b>			Pop 1	<b>4321</b>			Pop 1	<b>4321</b>		

(Continued)

TABLE 3 | Continued

Output ← Stack ← Input			Output ← Stack ← Input			Output ← Stack ← Input			Output ← Stack ← Input						
Start	-	-	<b>*3214</b>	Start	-	-	<b>*3241</b>	Start	-	-	<b>3421</b>	Start	-	-	<b>4321</b>
Push 3		3	214	Push 3		3	241	Push 3		3	421	Push 4		4	321
Pop 3	3		214	Pop 3	3		241	Push 4		4/3	21	Pop 4	4		321
Push 2	3	2	14	Push 2	3	2	41	Pop 4	4	3	21	Push 3	4	3	21
Pop 2	32		14	Push 4	3	4/2	1	Pop 3	43		21	Pop 3	43		21
Push 1	32	1	4	Pop 4	34	2	1	Push 2	43	2	1	Push 2	43	2	1
Push 4	32	4/1		Pop 2	342		1	Pop 2	432		1	Pop 2	432		1
Pop 4	324	1		Push 1	342	1		Push 1	432	1		Push 1	432	1	
Pop 1	<b>*3241</b>			Pop 1	<b>*3421</b>			Pop 1	<b>4321</b>			Pop 1	<b>4321</b>		

All and only the 213-avoiding orders, corresponding to attested DP orders (Cinque, 2005), are sorted into 4321. Input word order (top right) and output interpretation order (bottom left) are in bold.

### STACK-SORTING: LINEARIZATION, DISPLACEMENT, COMPOSITION, AND LABELED BRACKETS

In this section, I show that stack-sorting effectively encompasses linearization, displacement, and composition, as well as assigning brackets, labeled unambiguously and without search. Moreover, it does all of this without language-particular parameters.

In the standard (“Y-model”) view, linearization and composition are distinct interface operations, interpreting structures built in an autonomous syntactic module by Merge. In ULTRA, linearization goes in the other direction, loading surface word order item-by-item into memory, and reassembling it in order of compositional interpretation.

### Displacement Is a Natural Property of a Stack-Sorting Grammar

Displacement is a natural feature of stack-sorting; from one point of view, it is the basic property of the system. In standard accounts, constituents that compose together in the interpretation should appear adjacent in surface order. This arrangement is forced by phrase structure grammars. Displacement, whereby elements that compose together are separated by intervening elements in surface order, has always seemed a surprising property, in need of explanation.

Things work quite differently in ULTRA. A key assumption of the Merge-based view is discarded: there is no level of representation encompassing word order and the fseq within a unified higher-order object. Instead, word order and base hierarchy are disconnected sequences, related dynamically. Non-adjacent input elements can perfectly well end up adjacent in the output. Displacement, rather than being the exception, is the rule; every element in the surface order is “transformed,” passing through memory before retrieval for interpretation<sup>7</sup>.

have recently explored linguistic applications of transduction grammars, in the context of inter-language translation.

<sup>7</sup>Displacement under stack-sorting is limited to word order permutation within a single cycle. Long-distance displacement, such as successive-cyclic wh-movement, requires different mechanisms; see section Primacy vs. recency and the Duality of Semantics.

### Brackets and Labels without Primitive Constituency

The algorithm (4) implicitly assigns labeled bracketed structure<sup>8</sup> to each surface order, matching almost exactly the structures assigned by accounts like Cinque (2005). Explicitly, pushing (storage from word order to stack) corresponds to a left bracket, and popping (retrieval from stack for interpretation) to a right bracket. These operations apply to one element at a time; it is natural to think of that element as labeling the relevant bracket. See Table 4, which provides the stack-sorting computations for all surface permutations of a 3-element base.

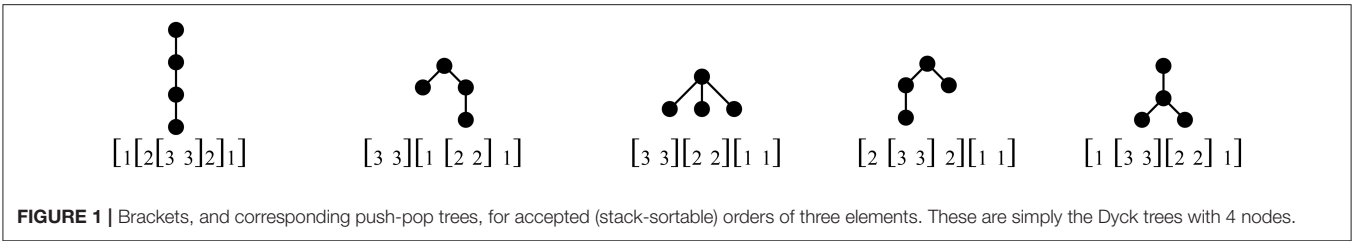
Examining these brackets, the sequence of pushes and pops (storage and retrieval) for each order implicitly defines a tree, as shown in Figure 1. These are the so-called Dyck trees<sup>9</sup>, the set of all ordered rooted trees with a fixed number of nodes (here, 4). Compare these to the binary-branching trees assigned under Cinque’s (2005) account, with non-remnant, leftward movement affecting a right-branching base (Figure 2). The brackets are nearly identical, as are their labels, taking some liberties with the technical details of Cinque’s account<sup>10</sup>.

Setting aside the 321 tree(s) for the moment, the Dyck trees are systematic, loss-less compressions of Cinque’s trees, with every subtree that is a right-branching comb in the Cinque tree replaced with a linear tree (see Jayaseelan, 2008) in the Dyck tree. For this correspondence, which amounts to pruning all terminals

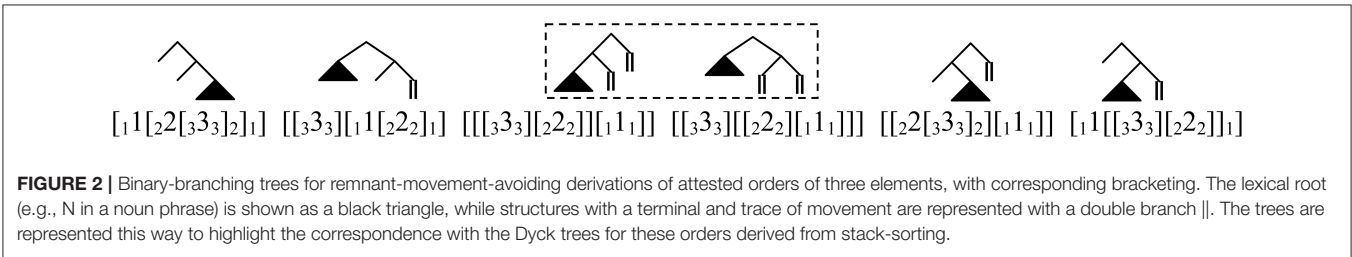
<sup>8</sup>Stack-sorting is intended as a parsing algorithm. There are standard techniques for extracting bracketed structure from strings with a stack-based parser, such as SR (shift-reduce) parsing. An SR parser has a set of “grammar rules,” specifying licensed surface configurations; when a set of elements on top of the stack match a grammar rule, they may be reduced, replacing them in the stack with the non-terminal symbol from the left-hand side of the rule (e.g., VP, NP on top of the stack may be reduced to S, by the rule S → NP VP). A sentence is successfully parsed if fully reduced to the start symbol S; reduction steps realize its phrase-structural analysis. This is quite unlike the stack-sorting procedure, which deploys no grammar rules, nor reduce steps, and applies parsing steps to one element at a time.

<sup>9</sup>The Dyck trees of successive sizes are counted by the Catalan numbers (1, 2, 5, 14, 42, ...). These numbers also count permutations avoiding any three-element subsequence.

<sup>10</sup>Technically, in Cinque’s theory the dependent modifiers do not label the phrases containing them. Instead, in line with Kayne (1994) Linear Correspondence Axiom, they are phrasal specifiers of silent functional heads. The labeling on the brackets derived instead more closely matches Starke’s representations.



**FIGURE 1** | Brackets, and corresponding push-pop trees, for accepted (stack-sortable) orders of three elements. These are simply the Dyck trees with 4 nodes.

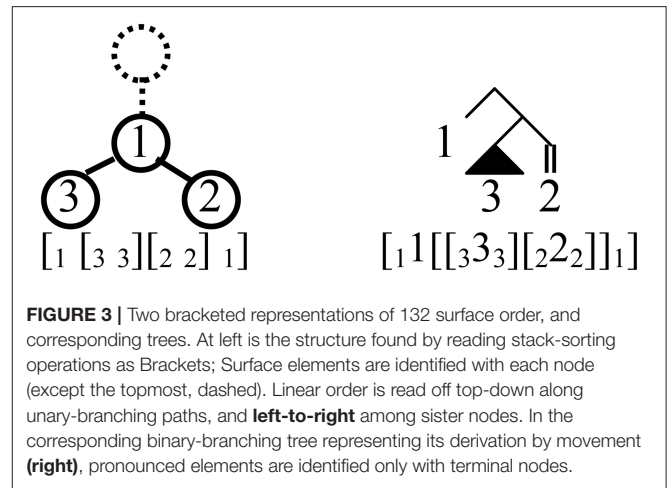


**FIGURE 2** | Binary-branching trees for remnant-movement-avoiding derivations of attested orders of three elements, with corresponding bracketing. The lexical root (e.g., N in a noun phrase) is shown as a black triangle, while structures with a terminal and trace of movement are represented with a double branch ||. The trees are represented this way to highlight the correspondence with the Dyck trees for these orders derived from stack-sorting.

in the binary tree, the lexical root (e.g., noun in a DP) must not be pruned. Elements from the surface order are associated to each node of the Dyck tree except the highest<sup>11</sup>, with linear order read left-to-right among sister nodes, and top-down along unary-branching paths. For example, for surface order 132, 1 is associated to the sole binary-branching node in its Dyck tree, 3 and 2 to its left and right daughters (Figure 3).

Meanwhile, 321 order, assigned a ternary tree by stack-sorting, has two remnant-movement-avoiding derivations.<sup>12</sup> In one possible derivation, 3 inverts with 2 immediately after 2 is Merged, then the 32 complex moves past 1 after 1 is Merged. In the other possible derivation, the full base structure is Merged first, then 23 moves to the left, followed by leftward movement of just 3<sup>13</sup>.

A key empirical question is whether 321 orders exhibit two distinct bracketed structures, as binary-branching treatments allow, or only the single, “flat” structure predicted here. The issue is even more acute for 4 elements, as in Universal 20, where there are up to 5 distinct Merge derivations<sup>14</sup> for 4321 order. Luckily, this (*N Adj Num Dem*) is the most common noun phrase order; future research should illuminate the issue<sup>15</sup>.



**FIGURE 3** | Two bracketed representations of 132 surface order, and corresponding trees. At left is the structure found by reading stack-sorting operations as Brackets; Surface elements are identified with each node (except the topmost, dashed). Linear order is read off top-down along unary-branching paths, and **left-to-right** among sister nodes. In the corresponding binary-branching tree representing its derivation by movement (**right**), pronounced elements are identified only with terminal nodes.

### Section Summary

Stack-sorting captures a surprising amount of syntactic machinery, normally divided among different modules. In the usual view, an autonomous generative engine builds constituent structures, interpreted at the interfaces by further processes of linearization and composition. In ULTRA, linearization and composition reflect a single procedure. Constituent structure is not primitive, but records the storage and retrieval steps by which stack-sorting assembles the interpretation<sup>16</sup>. This produces a bracketed surface structure, labeled appropriately, largely identical to the bracketed structure in accounts postulating movement (Internal Merge) from a uniform base (formed by External Merge). However, where standard theories countenance

<sup>11</sup>This departs from the usual view that words are terminals, with non-terminals representing constituents.

<sup>12</sup>Beyond collapsing ambiguous binary branching to flat, beyond-binary structure, the ternary Dyck tree for 321 order otherwise corresponds to the binary trees as indicated above: prune all terminals in the binary tree, preserving the lexical root (N).

<sup>13</sup>Some might object to extraction from already-moved objects, violating “Freezing.” However, such subextraction is required to derive attested *N-Dem-Adj-Num* (4132) order.

<sup>14</sup>For Cinque (2005), dedicated Agreement Phrases above each modifier-introducing category provide the landing sites of movement. This sharply reduces possible movements. But these AgrPs are technical devices introduced to comply with Kayne’s LCA, rather than a central part of his theory. See Abels and Neeleman (2012) for discussion.

<sup>15</sup>Cinque (2005, p. 320) gives the following partial list of languages with this order: Cambodian, Javanese, Karen, Khmu, Palaung, Shan, Thai, Enga, Dagaare, Ewe, Gungbe, Labu and Ponapean, Mao Naga, Selepet, Yoruba, West Greenlandic, Amele, Igbo, Kusaean, Manam, Fa d’Ambu, Nubi, Kugu Nganhcara, Cabécar, Kunama, and Maori.

<sup>16</sup>The claim that surface structure is an epiphenomenon of processing echoes ideas of Steedman’s Combinatory Categorical Grammar (CCG). He argues against viewing “[...] Surface Structure as a level of representation at all, rather than viewing it (as computational linguists tend to) as no more than a trace of the algorithm that delivers the representation that we are really interested in, namely the interpretation.” (Steedman, 2000, p. 3).

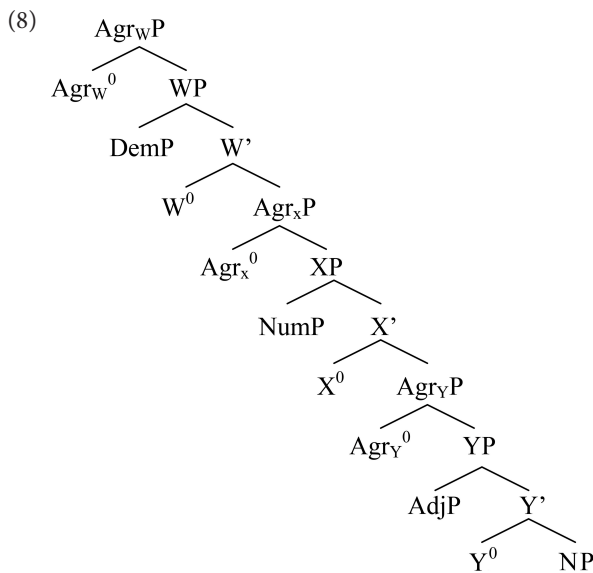
multiple derivations for some surface orders (and ambiguous binary-branching structure), the present account assigns unique beyond-binary bracketing. Significantly, there is no role for language-particular features to drive movement. Displacement is handled automatically by stack-sorting, and is in fact its core feature.

### COMPARISON WITH EXISTING ACCOUNTS OF UNIVERSAL 20

This section compares the stack-sorting account of Universal 20 to existing Merge-based accounts (Cinque, 2005; Steddy and Samek-Lodovici, 2011; Abels and Neeleman, 2012). I argue that the stack-sorting account is simpler, while avoiding problems that arise in each of these existing alternatives.

#### The Account of Cinque (2005)

Cinque proposes a cross-linguistically uniform base hierarchy, reflecting a fixed order of External Merge. He proposes that movement (Internal Merge) is uniformly leftward, while the base is right-branching, in line with Kayne’s (1994) LCA. He stipulates that remnant movement in the noun phrase is barred: each movement affects the noun, or a constituent containing it. His base structure for the noun phrase is (8).

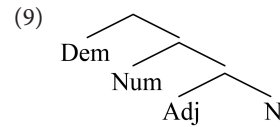


The overt modifiers are specifiers of dedicated functional heads (e.g., X<sup>0</sup>), below agreement phrases providing landing sites for movement. This structure, and his assumptions about movement, derives all and only the attested orders. The English-like order *Dem-Num-Adj-N* surfaces without movement; all other orders involve some sequence of movements of NP, or something containing it.

#### The Account of Abels and Neeleman (2012)

Abels and Neeleman (2012) modify Cinque’s analysis, discarding elements introduced to conform to the LCA (including agreement phrases and dedicated functional heads). They argue that the LCA plays no explanatory role; all that is required is that

movement is leftward, and remnant movement is barred. They allow free linearization of sister nodes, utilizing a considerably simpler base structure (9). They omit labels for non-terminal nodes as irrelevant to their analysis (Abels and Neeleman, 2012: 34).



In their theory, eight attested orders can be derived without movement, by varying the linear order of sisters. The remaining attested orders require leftward, non-remnant movement. In principle, their system allows a superset of Cinque’s (2005) derivations; some orders can be derived through linearization choices or through movement. However, restricting attention to strictly necessary operations, and supposing that free linearization is simpler than movement, their derivations are generally simpler than Cinque’s.

#### The Account of Steddy and Samek-Lodovici (2011)

Steddy and Samek-Lodovici (2011) offer another variation on Cinque’s (2005) analysis. They propose an optimality-theoretic account, retaining Cinque’s base structure (8). Linear order is governed by a set of Align-Left constraints (10), one for each overt element.

- (10)
- a. N-L – Align(NP, L, AgrWP, L)  
Align NP’s left edge with AgrWP’s left edge.
  - b. A-L – Align(AP, L, AgrWP, L)  
Align AP’s left edge with AgrWP’s left edge.
  - c. NUM-L – Align(NumP, L, AgrWP, L)  
Align NumP’s left edge with AgrWP’s left edge.
  - d. DEM-L – Align(DemP, L, AgrWP, L)  
Align DemP’s left edge with AgrWP’s left edge.
- (From Steddy and Samek-Lodovici, 2011: 450).

These alignment constraints incur a violation for each overt element or trace separating the relevant item from the left edge of the domain, and are variably ranked across languages. Attested orders are optimal candidates under some constraint ranking. The unattested orders are ruled out because they are “harmonic-bounded”: some other candidate incurs fewer higher-ranked violations, under any constraint ranking. Therefore, they can discard the constraints on movement that Cinque (2005) and Abels and Neeleman (2012) adopt. The leftward, non-remnant character of movement instead falls out from alignment principles.

#### Problems with Existing Accounts

Although these accounts differ in details, they share some problematic features. First, all of them capture the word order pattern in three tiers of explanation: (i) a uniform base structure, (ii) syntactic movement, and (iii) principles of linearization. In all three accounts, (i) describes the order of External Merge. Details of (ii) and (iii) vary between the accounts. For Cinque



**TABLE 4 |** Stack-sorting computations for orders of 3 elements.

	Output	←	Stack	←	Input
Start	-		-		<b>123</b>
[ Push			1		23
[ Push			2/1		3
[ Push			3/2/1		
] Pop	3		2/1		
] Pop	32		1		
] Pop	<b>321</b>				
	Output	←	Stack	←	Input
Start	-		-		<b>132</b>
[ Push			1		32
[ Push			3/1		2
[ Pop	3		1		2
] Pop	3		2/1		1
[ Push	32				1
] Pop	<b>321</b>				
	Output	←	Stack	←	Input
Start	-		-		<b>213</b>
[ Push			2		13
] Pop	2				13
[ Push	2		1		3
[ Push	2		3/1		
] Pop	23		1		
] Pop	<b>*231</b>				<b>FAILED SORT</b>

	Output	←	Stack	←	Input
Start	-		-		<b>231</b>
[ Push		2	31		
[ Push		3/2	1		
[ Pop	3	2	1		
] Pop	32		1		
[ Push	32	1			
] Pop	<b>321</b>				
	Output	←	Stack	←	Input
Start	-		-		<b>312</b>
[ Push			3		12
] Pop	3				12
[ Push	3		1		2
[ Push	3		2/1		
] Pop	32				1
] Pop	<b>321</b>				
	Output	←	Stack	←	Input
Start	-		-		<b>312</b>
[ Push			3		21
] Pop	3				21
[ Push	3		2		1
] Pop	32				1
[ Push	32		1		
] Pop	<b>321</b>				

Each order induces a unique sequence of pushes and pops, annotated with left or right brackets, respectively. The surface order is at topright within each computation, passing sequentially though memory to the output, at bottom left.

(2005) and Steddy and Samek-Lodovici (2011), all orders except *Dem-Num-Adj-N* involve movement; Abels and Neeleman (2012) require movement for only six attested orders. With respect to linearization, Cinque (2005) utilizes Kayne’s (1994) LCA; Abels and Neeleman (2012) have movement uniformly to the left, but base-generated sisters freely linearized on a language-particular basis; Steddy and Samek-Lodovici (2011) have language-particular constraint rankings.

These accounts all require different grammars for different orders. In Cinque (2005) system, features driving particular movements must be learned. The same is true for Abels and Neeleman (2012), with additional learning of order for sister nodes. Steddy and Samek-Lodovici (2011) require learning of the constraint ranking that gives rise to each order. All these accounts face trouble, therefore, with languages permitting freedom of order in the DP; in effect they must allow for underspecified or competing grammars, to capture the different orders.

Finally, all these accounts have some measure of structural or grammatical ambiguity, for some orders. For Cinque (2005), one kind of ambiguity comes about in choosing whether to move a functional category, or the Agreement phrase embedding it; this choice has no overt reflex. Although his theory sharply limits the number and landing site of

possible movements, these limitations are somewhat artificial; little substantive would change if we postulated further silent functional layers to host further movements, or allowed multiple specifiers. In the limit, this allows the full range of ambiguous derivations discussed in section Stack-Sorting: Linearization, Displacement, Composition, and Labeled Brackets. Abels and Neeleman’s (2012) approach allows this ambiguity among different movement derivations, as well as the derivation of many orders through either movement or reordering of sister nodes. Finally, Steddy and Samek-Lodovici (2011) face a different ambiguity problem: some orders are consistent with multiple constraint rankings (thus, multiple grammars).

### Comparison with the Stack-Sorting Account

The stack-sorting account fares better with respect to these issues. Instead of postulating separate tiers of base, movement, and linearization principles, the relevant machinery is realized in one algorithmic process. The sorting algorithm is universal, eschewing language-particular features to drive movement, order sister nodes, or rank alignment constraints. Such a theory is ideally situated to account for the free word order

phenomenon<sup>17</sup>. Furthermore, each order induces a unique sequence of storage and retrieval operations, tracing a unique bracketing. Within domains characterized by neutral word order and a single fseq, there is no spurious structural or grammatical ambiguity, for any word order.

## UNIVERSAL GRAMMAR AS UNIVERSAL PARSER

This section develops the view that stack-sorting can form the basis for an invariant performance mechanism, realizing Universal Grammar as a universal parser. This modifies traditional conclusions about competence and performance, while providing a novel view of what a grammar is.

### Rethinking Competence and Performance

In generative accounts, a fundamental division exists between competence and performance (Chomsky, 1965). Competence encompasses knowledge of language, conceived of as an abstract computation determining the structural decomposition of infinitely many sentences. Separate performance systems access the competence system's knowledge during real-time processing. In terms of Marr (1982) three-tiered description for information-processing systems, competence corresponds to the highest, computational level, specifying *what* the system is doing, and *why*. Performance corresponds, rather loosely, to the lower, algorithmic level, describing *how* the computation is carried out, step-by-step.

Of course, Marr's hierarchy applies to the information-processing in language, under the present theory as well as any other. However, the division of labor between these components is significantly redrawn here, with much more of the burden of explanation carried by performance<sup>18</sup>. A crucial difference is that in ULTRA, bracketed structure is not within the purview of competence. Instead, such structure arises in the interaction of competence with the stack-sorting algorithm, during real-time parsing. The knowledge ascribed to the competence grammar is simpler, including the innate fseq as a core component<sup>19</sup>. In a way, this aligns with the views of Chomsky's recent work, in which competence is fundamentally oriented for computing interpretations, with externalization "ancillary."

### A Universal Parser

A novel claim of ULTRA is that there is a single parser for all languages. This departs from the nearly universal assumption that parsers interpret language-particular grammars. But even within that traditional view, the appeal of universal mechanisms has been recognized.

<sup>17</sup>What requires explanation, from this point of view, is why languages should settle on distinct, relatively rigid word orders. See section Possible Extensions to a More Complete Theory of Syntax.

<sup>18</sup>Marcus has endorsed this mode of explanation: "[A] theory of parsing should attempt to capture wherever possible the sorts of generalizations that linguistic competence theories capture; there is no reason in principle why these generalizations should not be expressible in processing terms" (Marcus 1980 p. 10).

<sup>19</sup>See Chesi and Moro (2015) for related discussion, and a different perspective.

"The key point to be made, however, is that the search should be a search for universals, even—and perhaps especially—in the processing domain. For it would seem that the strongest parsing theory is one which says that the grammar interpreter itself is a universal mechanism, i.e. that there is one highly constrained grammar interpreter which is the appropriate machine for parsing all natural languages" (Marcus, 1980 p. 11).

The idea that "the parser is the grammar" has a long history; see Phillips (1996, 2003), Kempson et al. (2001), and the articles in Fodor and Fernandez (2015) for recent perspective. Fodor refers to this as the *performance grammar only* (PGO) view.

"PGO theory enters the game with one powerful advantage: there must be psychological mechanisms for speaking and understanding, and simplicity considerations thus put the burden of proof on anyone who would claim that there is more than this" (Fodor, 1978 p. 470).

However, while granting that this entails a simpler theory, Fodor rejects the idea, finding no motivation for movement outside an autonomous grammar (*ibid.*, 472). This presupposes that movement is fundamentally difficult for parsing mechanisms (which should prefer phrase-structural mechanisms to transformational ones). However, in ULTRA, displacement is not a complication over more basic mechanisms; displacement *is* the basic mechanism.

### Displacement Is Not Unique to Human Language

It is often said that displacement is unique to human language, and artificial codes avoid this property<sup>20</sup>. But displacement appears in coding languages, in exactly the same sense that it appears in ULTRA. A simple example illustrates: the order in which users press keys on a calculator is not the order in which the corresponding computations are carried out. In practice, calculators compile input into Reverse Polish Notation for machine use, via Dijkstra's Shunting Yard Algorithm (SYA).

The example is not an idle one; the stack-sorting algorithm (4) is essentially identical to the SYA<sup>21</sup>. Lexical heads (nouns and verbs) are "shunted" directly to interpretation, as numerical constants are in a calculator. Meanwhile the satellites forming their extended projections are stack-sorted according to their relative rank, just like arithmetic operators. In this analogy, cartographic ordering parallels the precedence order of arithmetic operators.

In fact, though the property is little used, the SYA is a sorting protocol; many input orders lead to the same internal calculation. As calculator users, we utilize one input scheme (infix notation), but others would do as well. The standardized input order for calculators has the same status as particular languages with respect to ULTRA: users may fall into narrow ordering habits, but the algorithm automatically processes many other orders.

<sup>20</sup>For example: "These 'displacement' properties are one central syntactic respect in which natural languages differ from the symbolic systems devised for one or another purpose, sometimes called 'languages' by metaphoric extension (formal languages, programming languages); there are other respects, including semantic differences" (Chomsky, 1995, p. 222).

<sup>21</sup>Thanks to Michael Jarrett for discussion.

## Grammaticality and Ungrammaticality

One of the central tasks ascribed to grammars is distinguishing grammatical sentences of a language from ungrammatical strings. In ULTRA, knowledge of grammaticality is very different from knowledge of ungrammaticality. The former kind of knowledge is fundamentally about computing interpretations. But the invariant process interpreting one language's surface order can equally interpret the orders of other languages. From this point of view, there is only one I-language, and a single performance grammar that delivers it. While this conclusion is appealing, an important question remains: where do individual languages come from, with apparently different grammars?

## POSSIBLE EXTENSIONS TO A MORE COMPLETE THEORY OF SYNTAX

This section addresses two kinds of problems that follow from interpreting stack-sorting as a performance device. The first concerns reconciling the theory with what is known about real-time language processing; the second concerns extending the model to properties of syntax that are left unexplained. Even discussing these problems in depth, much less justifying any solutions, is beyond the scope of this paper. The intent is merely to sketch the challenges, and indicate directions for further work.

### Reaction vs. Prediction: Incrementality and Rigid Word Order

With respect to processing, one problem is that this approach seems to be contradicted by strong evidence for word-by-word incrementality in comprehension (especially in the Visual World paradigm; see Tanenhaus and Trueswell, 2006). ULTRA is "pedestrian" in the sense Stabler (1991) cautions against. Within each domain, bottom-up interpretation cannot begin until the lexical root of the fseq is encountered.

One possibility for reconciling ULTRA with incrementality draws on the distinction between reactive processes, such as the stack-sorting procedure, and predictive processing (see Braver et al., 2007; Huettig and Mani, 2016). The idea is that stack-sorting is a reactive mechanism for language perception; this is contrasted with—and necessarily supplemented by—predictive capacities, associated with top-down processing, and production<sup>22</sup>. The latter system alone contains learned, language-particular grammatical knowledge. This proposal echoes other approaches with a two-stage parsing process, such as Frazier and Fodor's (1978) Sausage Machine. ULTRA resembles their Preliminary Phrase Packager (PPP), a fast low-level structure-builder, distinguished from the larger-scale problem-solving of their Sentence Structure Supervisor (SSS). Marcus expresses a similar view, describing a parser as a "fast, 'stupid' black box" (Marcus, 1980: 204) producing partial analyses, supplemented with intelligent problem-solving for building large-scale structure.

I suggest that evidence for word-by-word incrementality can be reconciled with the present theory through an interaction

<sup>22</sup>The so-called "P-chain" closely identifies prediction and production (see, e.g., Dell and Chang, 2014).

between reaction and prediction, exploiting the notion of "hyperactivity" (Momma et al., 2015). The idea is that comprehension can skip ahead, giving the appearance of incrementality, if a lexical root (noun or verb) is provided in advance by prediction. Something like this seems to be true.

"There is growing evidence that comprehenders often build structural positions in their parses before encountering the words in the input that phonologically realize those positions [...] To take just one example, in a head-final language such as Japanese it may be necessary for the structure building system to create a position for the head of a phrase before it has completed the arguments and adjuncts that precede the head" (Phillips and Lewis, 2013 p. 19).

A complementary predictive system could help solve two further problems for ULTRA: explaining how production is possible, and why there are distinct languages with different, relatively rigid word orders. The stack-sorting algorithm is a unidirectional parser; there is no trivial way of "reversing the flow" for production. Facing this uncertainty, it would be natural to rely on prediction to supply word order in production<sup>23</sup>. To simplify production, it is helpful for word order to be predictable; in turn, word order tendencies in the linguistic environment can be learned by this system. This suggests a feedback loop, and a plausible route for the emergence and divergence of relatively rigid word orders.

### Primacy vs. Recency and the Duality of Semantics

A number of important syntactic properties remain unexplained. In order to extend the proposal to a remotely adequate theory, these properties must be addressed somehow. These include, first, a cluster of syntactic properties relating to A-bar syntax, and the so-called Duality of Semantics. I suggest that this distinctive kind of syntax relates to an important distinction in short-term memory, between primacy and recency, drawing on Henson's (1998) Start-End Model (SEM)<sup>24</sup>. In Henson's model, primacy and recency are distinct effects, reflecting content-addressable coding of two aspects of serial position.

Recency is naturally associated with stack (last-in, first-out) memory. Primacy, on the other hand, is naturally described by queue (first-in, first-out) memory. Besides optimal order of access, there is another important difference between primacy and recency effects. Put simply, the first element in a sequence remains the first element, no matter how many more elements follow; the primacy signature of a given element is relatively stable over the time scale relevant to parsing. Recency is different: each element in a sequence is a new right edge, suppressing the accessibility to recency-based memory of everything that precedes it. Thus, we expect a kind of "use-it-or-lose-it" pressure within recency memory, but not primacy memory.

<sup>23</sup>The Dynamic Syntax framework (Kempson et al., 2001) adopts a broadly similar view of production as parasitic on comprehension (thanks to Colin Phillips for discussion).

<sup>24</sup>In discussing memory architectures for language processing, Caplan and Waters (2013) point out that SEM is "reasonably well-established" in the psychological literature as a model of short-term memory, and yet no existing theories of linguistic parsing incorporate it.

Tentatively, I would like to suggest that distinct primacy and recency memory codes underlie the Duality of Semantics, and the division between A-bar and A-syntax. Recency, associated with a stack, is the basis for information-neutral, local permutation, generally characterized by nesting dependencies<sup>25</sup>. Supposing that primacy is crucially involved in non-neutral, A-bar-like syntax suggests an account for a cluster of surprising properties. Most obvious is the association of discourse-information effects with the “left periphery”: the left edge of domains is where we expect primacy memory to play a significant role<sup>26</sup>. An involvement of primacy memory also suggests an analysis of Superiority effects in multiple wh-movement constructions. In Merge-based theories, such constructions (exhibiting crossing dependencies) are problematic, and require stipulative devices like Richards (1997) “Tucking-In” derivations. Thinking of the effects as involving primacy memory suggests a simpler account: ordering of multiple wh-phrases is a matter of first-in, first-out access (queue memory). A final property of this alternative syntactic system that can be rationalized is long-distance movement. Possibly, the availability of long-distance movement for A-bar relations results from the stability of primacy memory, making items encountered in the left periphery accessible for recall later without great difficulty, in contrast to recency memory (which can only support short, local recall). While this is suggestive, addressing the vast literature on A-bar syntax must be left to future research.

## What about Recursion?

A final problem looming in the background is recursion. ULTRA operates within syntactic domains characterized by a single fseq. This requires some comment, as recursion is a fundamental property of syntax. For recursion as well, properties of memory, and intervention of a complementary predictive system, might be crucial. Intriguingly, human episodic memory appears to be independently hierarchical in structure, perhaps unlike related animals (Tulving, 1999; Corballis, 2009). In the SEM model, episodic tokens are created for groups, within grouped sequences (Henson, 1998). Linguistic recursion requires some further mechanism for treating the group token corresponding to one sequence as an item token in another sequence.

As discussed in section Comparison with Existing Accounts of Universal 20, in ULTRA, structural ambiguity does not arise without ambiguity of meaning, within single domains. However, structural ambiguity arises inevitably when multiple domains are present, in terms of which domain embeds in

<sup>25</sup>Stack-sorting alone can handle some local crossing dependencies. This is surprising, given the usual identification of automata utilizing push-down stacks with context-free grammars, and nesting dependencies. For example, 1,423, 4,132, and 4,231 are admissible noun phrase orders (Dem-N-Num-Adj, N-Dem-Adj-Num, and N-Num-Adj-Dem). All three exhibit crossing relations, in that the (selectional) dependency between 4 and 3 crosses the dependency between 2 and 1. In Merge-based accounts, these orders require movement. But as these orders are 213-avoiding, they are stack-sortable.

<sup>26</sup>It may seem suspicious to associate A-bar relations of all kinds to the left periphery; what about wh-in situ constructions? Richards (2010) notes that in Japanese, in situ wh-phrases occur at the left edge of a special prosodic domain, which extends rightward to the complementizer where they take scope.

another, or where to attach an element that could discharge positions in two distinct domains. This is where the “fast, stupid black box” is helpless, and must call on other resources. One obvious source of help in stitching together multiple domains is a separate predictive system, with access to top-down knowledge of plausible meanings in context. The persistent problem of resolving embedding ambiguities also provides motivation for rigid word order, which sharply reduces attachment possibilities.

An important point is that brackets are defined relative to a particular fseq. Recursive embedding of one domain in another (for example, a nominal as argument of a verb) involves projection of a bracket corresponding to the entire embedded phrase, within the embedding domain<sup>27</sup>. Consider the following example.

(11) The dog chased a ball.

There are three sorting domains here: two nominal projections, embedded in a third, verbal projection (setting aside the possibility that clauses contain two domains, vP and CP phases). Their ULTRA bracketing appears below.

(12) NP1 = [the [dog dog] the]  
 NP2 = [a [ball ball] a]  
 VP = [NP1 [chased chased] [NP2 NP2] NP1]

This example illustrates the ambiguity that accompanies embedding. The issue is how to link the nominal phrases to positions in the verb’s fseq (i.e., to theta roles). As theta roles are not overtly expressed (case-marking is an unreliable guide), the reactive parser must draw on external means (for example, language-particular ordering habits, or predictions of plausible interpretations).

A final point about recursion returns to the issue of how calculators work, via Dijkstra’s Shunting Yard Algorithm. Such computations are recursive. But recursion isn’t handled by the parsing algorithm; rather, it arises at the level of interpretation, where partial outputs of arithmetic operations feed into further calculations. A similar conclusion (recursion is semantic, not syntactic) is possible within the present framework, given the similarities between ULTRA and the SYA. Notably, both

<sup>27</sup>Psycholinguistic evidence suggests that interpretation of clausal recursion proceeds top-down (Bach et al 1986, Joshi, 1990). Thus, in a recursive structure like [CP1 ... [CP2 ... [CP3 ... ]]], the order of interpretation is <CP1, CP2, CP3>. This suggests an intriguing extension of the present theory. Suppose recursive embedding is also parsed by stack-sorting. If the required output of stack-sorting recursive domains is 123-like (top-down) order among cycles, then we predict avoidance of 231-like orders. 231-avoidance is one way of expressing the Final-Over-Final Constraint (FOFC; see Sheehan et al., 2017 for a recent review). Thus, we can explain FOFC effects with this theory, insofar as they obtain over higher-order (recursive) structure. Consider, for example, one robust FOFC effect, the avoidance of V-O-Aux orders across Germanic (and beyond). The nominal is a distinct domain embedded within the clause. The clause itself arguably contains two cycles (vP and CP), with Aux in a higher cycle than V. Then V-O-Aux gives elements of vP, DP, CP, a 231-like order over the top-down embedding hierarchy [CP [vP [DP]]].

If this is on the right track, then 213- and 231-avoidance characterize different levels of structure (bottom-up assembly within local domains, vs. top-down recursion). I leave exploration of this possibility to future work.

procedures compile input into Reverse Polish Notation, a so-called concatenative programming language, expressing recursive hierarchical operations unambiguously in serial format.

## Do We Even Need an Algorithm?

I have shown how a particularly simple algorithm captures a range of syntactic phenomena. But the question is, why *this* algorithm? Other sorting procedures are possible in principle, and would lead to different permutation-avoidance profiles. How do we justify selecting stack-sorting as the right procedure for syntactic mapping?

There are three crucial ingredients. The first is the orientation of the system as a parser, mapping sound to meaning. This is not logically necessary; it is simply one of the reasonable choices. The second factor is the linear nature of sound and meaning. This is straightforward for sound sequences, but much less so for interpretations, where it is simply a bold hypothesis. The third ingredient is the choice of stack memory. This can plausibly be tied to the Modality effect: intelligible speech input engenders unusually strong recency effects (Surprenant et al., 1993). It seems a small leap to suppose that the formal stack employed in the algorithm may simply (and crudely) reflect the dominance of recency effects in memory for linguistic material.

So far, stack-sorting has been implemented with an explicit algorithm. That may be unnecessary. Rather than thinking of stack-sorting as a set of explicit instructions, we might reframe it as an *anti-conflict bias* between the accessibility of items in memory, in terms of recency effects, and retrieval for a rigid interpretation sequence. If that is on the right track, it is possible that no novel cognitive machinery had to evolve to explain these effects. What remains is to understand where the ordering of interpretations (the fseq) itself comes from, a matter on which I will not speculate here.

## CONCLUSION

Summarizing, a simple algorithm (4) maps 213-avoiding word orders to a bottom-up compositional sequence, while mapping 213-containing orders to deviant sequences. While the input and output of the mapping are sequences, hierarchical structure is present: the algorithmic steps realize left and right brackets, almost exactly where standard accounts place them. The account differs from standard accounts in assigning unambiguous bracketing to all orders.

This model improves on existing accounts of word order restrictions, which invoke additional stipulations (e.g., constraints on movement, together with principles of linearization), beyond core syntactic structure-building. In ULTRA, these effects fall out from a single real-time

process. In turn, syntactic displacement, long seen as a curious complication, emerges as the fundamental grammatical mechanism. No learning of language-particular properties is required; one grammar interprets many orders.

It should be clear that the system described here is only one part of syntactic cognition. This system builds one extended projection at a time; further mechanisms are required to embed one domain in another. However, that may be a virtue: it is tempting to identify the domains of operation for this architecture with phases, which are thus special for principled reasons.

Moreover, stack-sorting only handles information-neutral structure. This ignores another important component of syntax, so-called discourse-information structure, associated with potentially long-distance A-bar dependencies. This deficiency, too, may be a virtue, suggesting a principled basis for the Duality of Semantics. I speculated that primacy memory plays a central role in these effects, potentially explaining several curious properties (leftness, long distance, and crossing).

Raising our sights, the larger conclusion is that much of the machinery of syntactic cognition might reduce to effects not specific to language. Needless to say, this is just a programmatic sketch; future research will determine whether and how ULTRA's stack-sorting might be integrated into a more complete model of language.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## ACKNOWLEDGMENTS

Portions of this work were first developed in a graduate seminar at the University of Arizona (2015), and presented at the POL Symposium in Tokyo (2016), LCUGA 3 (2016), and the LSA Annual Meeting (2017). I am grateful to audiences at those venues, and to numerous colleagues for helpful discussion. I would like to single out the following for thanks: Klaus Abels, David Adger, Bob Berwick, Tom Bever, Andrew Carnie, Noam Chomsky, Guglielmo Cinque, Jennifer Culbertson, Sandiway Fong, Thomas Graf, John Hale, Heidi Harley, Norbert Hornstein, Michael Jarrett, Richard Kayne, Hilda Koopman, Diego Krivochen, Marco Kuhlmann, Massimo Piattelli-Palmarini, Colin Phillips, Paul Pietroski, Marcus Saers, Yosuke Sato, Dan Siddiqi, Dominique Sportiche, Juan Uriagereka, Elly van Gelderen, Andrew Wedel, and Stephen Wilson. I am especially grateful to the graduate students who participated in my Stack-Sorting Research Group (2015-2016). Of course, all errors and misunderstandings in this paper are my own.

## REFERENCES

- Abels, K., and Neeleman, A. (2012). Linear asymmetries and the LCA. *Syntax* 15, 25–74. doi: 10.1111/j.1467-9612.2011.00163.x
- Barbiers, S., van der Auwera, J., Bennis, H., Boef, E., de Vogelaer, G., and van der Ham, M. (2008). *Syntactic Atlas of the Dutch Dialects, Vol. II*, Amsterdam: Amsterdam University Press.
- Berwick, R., and Chomsky, N. (2016). *Why Only Us: Language and Evolution*. Cambridge, MA: MIT Press.
- Braver, T. S., Gray, J. R., and Burgess, G. C. (2007). “Explaining the many varieties of working memory variation: dual mechanisms of cognitive control,” in *Variation in Working Memory*, eds A. Conway, C. Jarrold, M. Kane, A. Miyake, and J. Towse (New York, NY: Oxford University Press), 76–106.
- Caplan, D., and Waters, G. (2013). Memory mechanisms supporting syntactic comprehension. *Psychon. Bull. Rev.* 20, 243–268. doi: 10.3758/s13423-012-0369-9
- Chesi, C., and Moro, A. (2015). “The subtle dependency between Competence and Performance,” in *50 Years Later: Reflections on Chomsky’s Aspects*, eds Á. J. Gallego and D. Ott (Cambridge, MA: MIT Working Papers in Linguistics), 33–46.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT press.
- Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry* 36, 1–22. doi: 10.1162/0024389052993655
- Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.
- Cinque, G. (2005). Deriving Greenberg’s universal 20 and its exceptions. *Linguistic Inquiry* 36, 315–332. doi: 10.1162/0024389054396917
- Corballis, M. C. (2009). Mental time travel and the shaping of language. *Exp. Brain Res.* 192, 553–560. doi: 10.1007/s00221-008-1491-9
- Dell, G. S., and Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Phil. Trans. R. Soc. B* 369:20120394. doi: 10.1098/rstb.2012.0394
- Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9, 427–473.
- Fodor, J. D., and Fernandez, E. (2015). Special issue on grammars and parsers: toward a unified theory of language knowledge and use. *J. Psycholinguist. Res.* 44, 1–5. doi: 10.1007/s10936-014-9333-3
- Frazier, L., and Fodor, J. D. (1978). The sausage machine: a new two-stage parsing model. *Cognition* 6, 291–325. doi: 10.1016/0010-0277(78)90002-1
- Greenberg, J. (1963). “Some universals of grammar with particular reference to the order of meaningful elements,” in *Universals of Language*, ed J. Greenberg (Cambridge, MA: MIT Press), 73–113.
- Henson, R. N. (1998). Short-term memory for serial order: the start-end model. *Cogn. Psychol.* 36, 73–137. doi: 10.1006/cogp.1998.0685
- Huetting, F., and Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Lang. Cogn. Neurosci.* 31, 19–31. doi: 10.1080/23273798.2015.1072223
- Jayaseelan, K. A. (2008). Bare phrase structure and specifier-less syntax. *Biolinguistics* 2, 87–106.
- Joshi, A. K. (1990). Processing crossed and nested dependencies: an automaton perspective on the psycholinguistic results. *Lang. Cogn. Process* 5, 1–27. doi: 10.1080/01690969008402095
- Kayne, R. (1994). *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.
- Kempson, R., Meyer-Viol, W., and Gabbay, D. (2001). *Dynamic Syntax*. Oxford: Blackwell.
- Knuth, D. (1968). *The Art of Computer Programming, Vol. 1: Fundamental Algorithms*. (Reading, MA: Addison-Wesley).
- Koopman, H. (2000). *The Syntax of Specifiers and Heads*. London: Routledge.
- Marcus, M. P. (1980). *Theory of Syntactic Recognition for Natural Languages*. Cambridge, MA: MIT press.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York, NY: W. H. Freeman.
- Momma, S., Slevc, L. R., and Phillips, C. (2015). “The timing of verb planning in active and passive sentence production,” in *Poster Presented at the 28th Annual CUNY Conference on Human Sentence Processing* (Los Angeles, CA).
- Phillips, C. (1996). *Order and Structure*. Doctoral dissertation, Massachusetts Institute of Technology.
- Phillips, C. (2003). Linear order and constituency. *Linguistic Inquiry* 34, 37–90. doi: 10.1162/002438903763255922
- Phillips, C., and Lewis, S. (2013). Derivational order in syntax: evidence and architectural consequences. *Stud. Ling.* 6, 11–47.
- Richards, N. (1997). *What Moves Where When in Which Language?* Doctoral dissertation, Massachusetts Institute of Technology.
- Richards, N. (2010). *Uttering Trees*. Cambridge, MA: MIT Press.
- Rizzi, L. (1997). “The fine structure of the left periphery,” in *Elements of Grammar: Handbook in Generative Syntax*, ed L. Haegeman (Dordrecht: Kluwer), 281–337.
- Schmid, T., and Vogel, R. (2004). Dialectal variation in German 3-verb clusters: a surface oriented OT account. *J. Compar. Germanic Ling.* 7, 235–274. doi: 10.1023/B:JCOM.0000016639.53619.94
- Sheehan, M., Biberauer, T., Roberts, I., Pesetsky, D., and Holmberg, A. (2017). *The Final-Over-Final Condition: A Syntactic Universal, Vol. 76*, Cambridge, MA: MIT Press.
- Stabler, E. P. Jr (1991). “Avoid the pedestrian’s paradox,” in *Principle-Based Parsing: Computation and Psycholinguistics*, eds R. Berwick, S. Abney, and C. Tenny (Dordrecht: Kluwer), 199–237.
- Starke, M. (2004). “On the inexistence of specifiers and the nature of heads,” in *The Cartography of Syntactic Structures, Vol. 3: Structures and Beyond*, ed A. Belletti. (New York, NY: Oxford University Press), 252–268.
- Steddy, S., and Samek-Lodovici, V. (2011). On the ungrammaticality of remnant movement in the derivation Greenberg’s universal 20. *Linguistic Inquiry* 42, 445–469. doi: 10.1162/LING\_a\_00053
- Steedman, M. (2000). *The Syntactic Process*. Cambridge, MA: MIT Press.
- Surprenant, A. M., Pitt, M. A., and Crowder, R. G. (1993). Auditory recency in immediate memory. *Q. J. Exp. Psychol.* 46, 193–223. doi: 10.1080/14640749308401044
- Tanenhaus, M. K., and Trueswell, J. C. (2006). “Eye movements and spoken language comprehension,” in *Handbook of Psycholinguistics*, eds M. Traxler and M. A. Gernsbacher (New York, NY: Elsevier Academic Press), 863–900.
- Tulving, E. (1999). “On the uniqueness of episodic memory,” in *Cognitive Neuroscience of Memory*, eds L. Nilsson and H. J. Markowitsch (Ashland, OH: Hogrefe and Huber), 11–42.
- Wurmbrand, S. (2006). “Verb clusters, verb raising, and restructuring,” in *The Blackwell Companion to Syntax, Vol. V*, eds M. Everaert and H. van Riemsdijk (Oxford: Blackwell), 227–341.
- Zwart, C. J. (2007). Some notes on the origin and distribution of the IPP-effect. *Groninger Arbeiten zur Germanistischen Linguistik* 45, 77–99.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Medeiros. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Assessing the Role of Experimental Evidence for *Interface* Judgment: Licensing of Negative Polarity Items, Scalar Readings, and Focus

Anastasia Giannakidou<sup>1</sup> and Urtzi Etxeberria<sup>2\*</sup>

<sup>1</sup> Department of Linguistics, University of Chicago, Chicago, IL, United States, <sup>2</sup> Centre National de la Recherche Scientifique, IKER (UMR 5478), Bayonne, France

## OPEN ACCESS

### Edited by:

Ángel J. Gallego,  
Universitat Autònoma de Barcelona,  
Spain

### Reviewed by:

Leticia Pablos,  
Leiden University Centre for  
Linguistics (LUCL), Netherlands  
John E. Drury,  
Stony Brook University, United States

### \*Correspondence:

Urtzi Etxeberria  
u.etxeberria@iker.cnrs.fr

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 08 May 2017

**Accepted:** 15 January 2018

**Published:** 21 February 2018

### Citation:

Giannakidou A and Etxeberria U  
(2018) Assessing the Role of  
Experimental Evidence for Interface  
Judgment: Licensing of Negative  
Polarity Items, Scalar Readings, and  
Focus. *Front. Psychol.* 9:59.  
doi: 10.3389/fpsyg.2018.00059

This paper reviews a series of experimental studies that address what we call “interface judgment,” which is the complex judgment involving integration from multiple levels of grammatical representation such as the syntax-semantics and prosody-semantics interface. We first discuss the results from the ERP literature connected to NPI licensing in different languages, paying particular attention to the N400 and the P600 as neural correlates of this specific phenomenon and focusing on the study by Xiang et al. (2016). The results of this study show evidence that there are two distinct NPI licensing mechanisms, i.e., licensing and rescuing, in line with Giannakidou (1998, 2006). Then we discuss an acceptability judgment task on Greek NPIs which supports the negativity as a scale hypothesis (Zwarts, 1995, 1996; Giannakidou, 1998). For the semantics-prosody interface judgment, we discuss two types of findings on two different phenomena and languages: (i) the study by Giannakidou and Yoon (2016) on scalar and non-scalar NPIs in Greek and Korean, which serves as the foundation for Chatzikonstantinou’s (2016) study of production data showing distinct prosodic properties in emphatic (scalar) and non-emphatic (non-scalar) Greek NPIs; (ii) a (production and perception) study by Etxeberria and Irurtzun (2015) on the prosodic disambiguation of the scalar/non-scalar readings of sentences containing the focus particle “ere” in Basque. The main conclusion of the paper is that experimental methods of the kind discussed in the paper are useful in establishing physical, quantitative correlates of interface judgment.

**Keywords:** interface judgment, negative polarity items, scalar items, FOCUS, prosody

## FRAMING THE TOPIC: MEANING AND INTERFACE JUDGMENT

What does it mean for speakers to have a linguistic judgment about meaning? How does the semantic judgment differ from the judgment about syntax, and how are the two to be distinguished from use errors or lexical failures? Are there hallmarks of syntax-semantics integration, and if so what are they? Are there grammatical phenomena that allow us to pinpoint physical correlates of this integration? Since sentences are typically *uttered*, what role does prosody play in disambiguating or bringing about additional dimensions of meaning that reflect different modules (semantics vs. pragmatics)? These are some of the questions that we discuss in the present article.

Let us define “interface judgment” as the judgment that comes from integrating representations from multiple grammatical levels such as, for instance, syntax and semantics, and semantics and prosody. The issues of interface judgment have *not* featured prominently in the older experimental literature, which tended to mostly study the nature of morpho-syntactic judgment. The reason is understandable: morphological and syntactic success or failure, such as e.g., person, gender, number agreement, or grammatical case inform us directly about what is generated (or not) by the grammar, and the judgment seems pretty robust. Since Chomsky (1957, 1964), the field has generally accepted that speakers’ reactions to, and intuitions about, “grammatical” and “ungrammatical” are relatively clear. Grammatical is a structure that is generated by the rules of the grammar; ungrammatical is one that cannot be generated by the grammar. Chomsky notes that speakers’ intuitions about grammaticality differ from intuitions about mere anomaly. Notice the contrast below:

- (1) Apples grow on trees (grammatical, meaningful).
- (2) \*Apples grew on trees (ungrammatical: ill-formed morphology of past tense).
- (3) #Apples grow on noses (grammatical, semantically anomalous).

Semantic anomaly is distinct from morphosyntactic failure (\*) and is marked with #, as indicated in (3). In (3), anomaly arises because of *lexical incompatibility*: the predicate *grow* doesn’t combine conceptually with “on noses” when it comes to apples; apples don’t grow on noses but on trees, in the US, in the fall, and the like. Warts, on the other hand, may indeed grow on noses. Likewise, (4) is odd because *Jason* is the name of a person, and a person cannot combine with the predicate “has a population of 3 million.”

- (4) Athens has a population of 3 million.
- (5) #Jason has a population of 3 million.

(4) is semantically well-formed while (5) is ill-formed due to the anomalous combination of subject and predicate. Sentences like (5) are known as category errors. We are dealing clearly with lexical mismatches that are nevertheless grammatical, i.e., generated by the grammar.

Other cases of semantic anomaly are produced with tautology and contradiction:

- (6) #Daniel Day-Lewis is Daniel Day-Lewis.
- (7) #It is raining outside and it is not raining outside.

These sentences are under-informative and therefore meaningless. A tautology is always true (analytic), and a contradictory sentence is always false; neither sentence, therefore, conveys information. Uninformative sentences are *infelicitous*—and often speakers try to make sense by enriching the meaning. For instance, (6) can be used to mean that Daniel Day-Lewis is a great actor, so even in a bad movie he will shine. Likewise (7) can be manipulated to mean that it is not *truly*

raining, i.e., it rains only a little bit. Speaker meaning is malleable this way as hearers strive to make sense of the infelicitous messages they get<sup>1</sup>.

The question is: is semantic judgment a felicity judgment, or is it more complex including interaction with morphosyntax, and perhaps prosody? This is not a trivial question to ask. Indeed, there are linguistic phenomena suggesting that semantic judgment is complex, and relies on *integrating* information from multiple levels of grammatical representation such as semantics, morphosyntax, and prosody. These three are central levels of grammatical representation, and studying their interfaces can be very useful in uncovering the nature of complex linguistic judgment. Two phenomena stand out as particularly illustrative cases: anaphoric pronouns and negative polarity items (NPIs), and we will study the latter here. Both involve syntax-semantics long distance dependencies in that the distribution of the anaphor and NPI are constrained because of a semantic requirement that forces a particular syntactic relation that is not local (i.e., it does not involve adjacent elements). It is therefore no surprise that there have been attempts to unify the two phenomena (Progovac, 1994). Giannakidou’s (1998) and Giannakidou and Quer’s (2013) concept of *dependent variable* (to be discussed later) can be understood as the semantic correlate of a syntactic anaphor.

The early literature (Ladusaw, 1980) defended a syntax-semantics view of NPIs as ungrammatical due to semantic and syntactic constraints; more recently, Giannakidou (1997, 1998), and Giannakidou and Quer (2013) develop this integration view further. Some literature, however, obliterates the difference between felicity and grammaticality, and treats NPI failures on a par with infelicities. This path had been taken in the early pragmatic scales tradition (Fauconnier, 1975; Israel, 1996, 2011), and has recently been pursued by Kadmon and Landman (1993), Krifka (1995), and to a certain extent by Chierchia (2006, 2013). In these approaches, NPI failures are characterized as “severe” infelicities, or “special” contradictions that lead to ungrammaticality (see also the free choice accounts in this spirit such as Aloni, 2007; Menéndez-Benito, 2010). The deciding criterion is not speaker’s intuition or a concrete psycholinguistic measure; typically, no deciding criterion is offered. Sometimes, authors refer to a post-compositional computation (suggested by Gajewski, 2002 in unpublished work) as a recipe for when to call a contradiction “grammatical,” and when to call it “ungrammatical.” However, unless we have some appeal to intuition or some other kind of evidence, it is difficult to see the distinction between a “grammatical” and “ungrammatical” contradiction as more than just highly speculative—a mere declaration, in fact, of the choice to avoid addressing the complexity in the nature of NPI licensing.

NPIs are words like *any*, *either*, *ever* which systematically fail when they do not occur in the scope of negation:

- (8) a. \*Bill brought any presents.

<sup>1</sup>Wellwood et al. (under review) discuss a similar malleability observed in comparative constructions like *More people have been to Russia than I have*. Such sentences tend to be perceived as meaningful by native speakers of English, but upon closer reflection, they are judged to be incoherent.



- b. Bill didn't buy any presents.
- c. \*Bill talked to John either.
- d. Bill didn't talk to John either.

Negation therefore “licenses” the NPI. The early literature (Klima, 1964; Ladusaw, 1980; Linebarger, 1987) typically marks NPI failures with \* and not #, reflecting the judgment that the NPI failure is not a mere use-based failure or lexical anomaly. Ladusaw proposed the concept of *semantic filtering*. Semantic filtering requires a module of grammar, just like binding theory (Chomsky, 1981), and this is a clear admission that NPI licensing is a grammatical and not a merely pragmatic phenomenon. The reason why NPIs fail is that they need to co-occur with a licenser, and the licenser has a specific semantic property. NPI-licensing, then, illustrates a synergy between semantics and syntax, and raises the question of well-formedness that is determined by both. In addition to *any*, *either*, we see indeed that many NPIs appear to be subject to severe grammatical constraints, e.g., Greek NPIs, “n-words” (see Laka, 1994; Giannakidou, 2006; Giannakidou and Zeijlstra, 2017). Generally, the literature on NPIs (which we will not review here, but see Giannakidou, 2011 for a recent overview) has shown that a substantial part of NPI violations involve grammatical violations that have to do with syntactic constraints too. Corpus studies such as Hoeksema (2010) on Dutch and English NPIs have also been instrumental in illustrating the (often severely) limited distributions of NPIs.

So, how does this semantico-syntactic integration judgment differ from infelicity, i.e., the merely pragmatic conflict of lexical anomaly and uninformativity? It is important to raise this question because sometimes, as we mentioned earlier, the difference is blurred. Based on intuition alone it is hard to make the case for different *types* of judgment—and here the contribution of experiments and quantitative data can be instrumental, we will argue.

In what follows, we will review literature that reveals the physical (neural) correlates of the “interface judgment,” which we take to be the judgment typical of interface phenomena that rely on integrating multiple levels of grammatical representation. We will study two prominent cases of interface judgment involving syntax-semantics and prosody-semantics integration. We will review (a) ERP literature showing distinct neural patterns for syntax-semantics integration with NPIs, (b) experiments, including mere acceptability judgment tasks, illustrating the usefulness of such methodology in extracting more refined sets of data, and (c) interaction of prosody with scalar structure and focus. Regarding focus, we will study recent work on the Basque focus particle *ere* “also/even.” Our goal is to show that experimental methodologies can be instrumental in revealing richer sets of data that make visible interactions between levels of grammatical representation—one can therefore be hopeful that experimental methodologies, at least of the kind discussed here, can further aid the understanding of syntax-semantics and prosody-semantics-pragmatics interface.

The promise, of course, does not come without open questions, especially since this is a new research paradigm with various methodologies. There certainly are many questions to be explored in terms of comparing the methodologies. However, it is impossible to offer a general comparison between methodologies in this short survey; our focus is rather the empirical phenomena of NPI-licensing, focus, and the particle *even*.

One might think that perhaps the notion of integration judgment is too broad, and that upon reflection it might be hard to think of language phenomena that do not, to some extent, require reference to multiple levels or interfaces (We thank John Drury for raising this point). While this concern is well taken, we believe it is overstated. There are plenty of morphosyntactic phenomena that do not make recourse to additional grammatical levels. For example, a failure to indicate the correct morphological tense in our earlier sentence (2) (\**growed*, instead of *grew*) does not require integration; nor do case assignment, and various kinds of agreement. For such phenomena, it is accepted that we don't need a level other than morphosyntactic computation, and we are not aware of any analyses that treat these phenomena as involving (non-trivial) integration. At the same time, it is true that integration phenomena are plentiful. If we can convince that experimental methodologies of the kind discussed here are helpful in teasing apart the various grammatical levels in NPI licensing and with focus, then we can be optimistic that such methodologies carry promise for other integration phenomena too<sup>2</sup>.

The paper develops as follows. In section NPI Licensing and the Syntax-Semantics Interface, we start with discussion of the interface nature of NPI licensing. We conclude that NPI licensing is a dependency that involves both semantics (matching, binding) and syntax (c-command). In section Interface Judgment with NPIs: Two Recurring Components, we discuss ERP evidence for the two components. In section Empirical Variation and the Scale of Negativity, we discuss an acceptability judgment study revealing a gradient status for NPI licensers in terms of negativity in Greek. In the first part of section Intonation and Meaning: Disambiguating Scalar Components, we discuss the role of intonation in Greek NPIs relying on recent work by Chatzikonstantinou (2016), who presents evidence that prosodic prominence in Greek NPIs indicates the presence of a scalar component. Finally, in the second part of section Intonation and Meaning: Disambiguating Scalar Components, we discuss the Basque additive particle *ere* “also/even,” shown in Etxeberria and Irurtzun (2015) to obtain scalar readings with prosodic prominence. We conclude in section General Conclusions.

<sup>2</sup>Crucially, it is not our goal to offer a comprehensive account of what might be meant by “integration” in terms of a finite set (or classes) of operations. The syntax semantics interface, for instance, and the prosody semantics interface will involve rather different integration operations—and some methodologies will likewise be better suited to capture certain integrations than others. It is definitely desirable to reach an account of operations that sets these apart in some interesting way from other kinds of processing, but that would require more extensive consideration of data than we can offer at present. We thank John Drury for raising this question.

## NPI LICENSING AND THE SYNTAX-SEMANTICS INTERFACE

Polarity is a pervasive phenomenon in natural language, and has received considerable attention since Klima's (1964) seminal work on English negation (see Giannakidou, 2011, 2017 for overviews and references). A "negative polarity item" (NPI), as we said earlier, is an expression that has limited distribution because it requires a preceding negation. Recall:

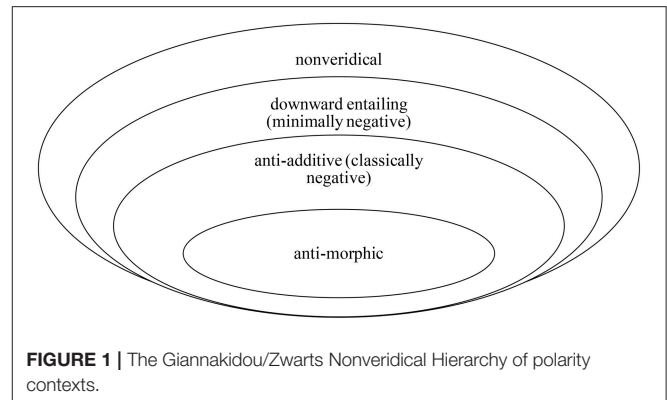
- (9) Nicholas didn't say anything.  
 (10) \*Nicholas said anything.  
 (11) Nicholas hasn't ever talked to Ariadne.  
 (12) \*Nicholas has ever talked to Ariadne.

*Anything* and *ever* are NPIs and need negation for grammaticality<sup>3</sup>; negation is said to license the NPI, it is thus the "licenser." An NPI-licenser can be non-negative but also nonveridical (Zwarts, 1995; Giannakidou, 1997, 1998 among others). NPIs therefore appear, in addition to negation, also in questions, with modal verbs, imperatives etc. We illustrate below with *any* and Greek NPIs:

- (13) a. Milises me kanenan?question  
 talked.2sg with anybody  
 Did you talk to anybody?  
 b. Patise kanena pliktro. imperative  
 press.imper any.NPI key  
 Press any key.  
 c. O Janis bori na milisi  
 the John may subj talk.3sg  
 me kanenan kathijiti (an theli)  
 with anybody professor (if wants.3sg)  
 John may talk to anybody (if he wants to).

Questions and modals are nonveridical, i.e., they don't entail the truth of  $\phi$ . Negation is also nonveridical: *not*  $\phi$  does not entail  $\phi$ . NPIs appear in nonveridical contexts, i.e., contexts in the scope of nonveridical operators. As summarized below, the negative is also nonveridical (Figure 1).

Nonveridicality includes negation as a subcase while allowing a wider distribution of NPIs in non-negative contexts. Downward entailment is the minimum of negativity (or, *minimal negation*, see Zwarts, 1995). Anti-additive and anti-morphic operators (*nobody*, *not*) are classically negative, which means they are stronger: they satisfy three (anti-additive) or all four (anti-morphic) of the de Morgan laws for complementation. Giannakidou and Zwarts, therefore, envision NPI licensers as having *gradient* strength: nonveridical non-negative licensers such as modals and questions are the weakest, negative licensers are stronger than non-negative, and within the negative class, minimal negation is weaker than classical negation. This predicts,



as is easy to see, differences in the licensing potential of expressions, and in section Empirical Variation and the Scale of Negativity we show that a simple acceptability judgment task can detect empirically the differences<sup>4</sup>.

In the present paper, we focus on the interaction of NPIs with negation. NPIs need negation to be licensed. Licensing, crucially, is both a semantic and a syntactic requirement: it is the semantic requirement that there be a negation in the sentence, and the syntactic requirement that the NPI *be in the scope* of negation, which translates into a need for the NPI to be c-commanded by negation:

- (14) a. Bill didn't see any student.  
 b. \*Any student didn't see Bill.  
 (15) a.  $\neg \exists x$ . student ( $x$ )  $\wedge$  see (Bill,  $x$ )  
 b.  $\exists x$ . student ( $x$ )  $\wedge$   $\neg$  see (Bill,  $x$ )

*Any student* can only be interpreted in the scope of negation—the scoping in (15b) where there is one person such that Bill didn't see her (but others that he saw), is impossible. Furthermore, appearance of *any* to the left of negation is generally prohibited (14b) (though there are NPIs that are not subject to this surface constraint, e.g., Greek NPIs and the Turkish NPIs considered in Yanilmaz and Drury, 2018). In the example below, we see more effects of the syntactic constraint on c-command:

- (16) a. \*The sister of no student said anything.  
 (17) b. No student said anything.

*No student* licenses *any* in (16) but not in (17) because, although *no student* linearly precedes the NPI, it does not c-command it. NPI licensing thus manifests a true semantics and syntax dependency, and presents a prime case for studying the interaction between these two levels. Giannakidou (1997) formulates licensing as follows:

- (18) a. *Licensing* (Giannakidou, 1997)  
 b.  $R(\beta, \alpha)$ ; where  $R$  is the scope relation;  $\alpha$  is the polarity item;  $\beta$  is a negative or nonveridical expression which serves as the licenser.

<sup>3</sup>*Any* can receive free choice readings in modal contexts (Giannakidou, 2001; Giannakidou and Cheng, 2006). We omit consideration of free choice here to keep things simple. Free choice is also a polarity phenomenon.

<sup>4</sup>Nonveridicality plays a decisive role in the selection of subjunctive and other non-assertive moods (see Giannakidou, 1995, 1998, 2009; Quer, 1998), hence it is a logical category with broader applications in grammar. Mood selection can certainly be understood as an interface phenomenon.

Licensing requires that the NPI  $\alpha$  be in the scope of  $\beta$ . R is a “scope” relation. Scope is both a semantic relation—a matching relation of polarity or sensitivity features (Giannakidou, 1997; Zeijlstra, 2004)—and the syntactic relation of *c*-command, or clause boundary (as it appears in Greek, Turkish, and Romance *n*-word NPIs)<sup>5</sup>.

Let us see now how NPI licensing is manifested in experimental data. Based on intuition alone, it is very hard to tease apart the semantic from the syntactic component of licensing.

## INTERFACE JUDGMENT WITH NPIS: TWO RECURRING COMPONENTS

NPIS have been studied in the past 15 years with the aid of event-related potentials (ERPs). This technique has been useful because it examines the temporal dynamics of syntactic, lexical semantic, and potentially syntactico-semantic dimensions of language processing—and the data for NPIS have been relatively consistent. ERP-profiles with NPIS have systematically included both LAN/P600 and N400 effects. The P600 is linked to syntactic processing and more broadly integration processes (as we elaborate later, see Kuperberg, 2013); N400 can be taken to index lexical semantic processing, or semantic matching (as argued recently in Xiang et al., 2016), in a way consistent with the view that NPI-licensing involves semantic-syntactic integration.

### N-400 and Lexical Matching

As the name suggests, the N400 is a negative-going waveform that peaks at approximately 400 ms, with a primarily centro-posterior scalp distribution. The amplitude of the N400 evoked by an incoming word indexes the degree to which that word’s semantic features match semantic features that have been pre-activated in the context at the time of encounter (Lau et al., 2008; Kutas and Federmeier, 2011; Kuperberg, 2013). The term “pre-activation” has often been associated with active prediction of specific lexical items, but Xiang et al. (2016) use it to refer more generally to the activation of relevant semantic features, regardless of whether active prediction or expectation of the upcoming word is at work.

Lau et al. (2008) offer a very lucid discussion of how the N400 can be used to reflect semantic effects related to “anomaly” and “expectation”—both relevant for NPI licensing as we saw at the beginning of the paper. The N400 response involves

the presentation of a congruent or incongruent word before a word target (such as “coffee–tea” or “chair–tea”). A “semantically supportive” context elicits a response of smaller amplitude in the 300–500 ms interval, and although the effects of sentential context on the N400 response may be bigger in magnitude, collectively, we refer to this modulation as the “N400 effect.” Lau et al. state further that the amplitude of the N400 response “is modulated not only by the degree of anomaly *per se*, but also by *predictability*. [emphasis ours]. A *less expected* sentence endings generate a *larger* N400 [emphases ours] response than highly expected ones, even when both endings are semantically congruent (for example, “I like my coffee with cream and honey” would generate a larger N400 response than “I like my coffee with cream and sugar”) (Lau et al., 2008 p. 921).

In the context of NPI licensing, the (negative or nonveridical) licenser establishes an expectancy that, given the above, predicts an N400 response on the NPI. Expectancy and predictability explain why lexical anomalies (“I like my coffee with cream and sugar/socks,” “Apples grow on noses”) typically show the effect, but it would be erroneous to state that N400 indexes merely semantic anomaly (as it has indeed been stated in the past). The N400 is expected to show up in more patterns of incongruence, as indeed is the case (see Lau et al. for data and specific references). For NPIS, the presence of the N400 will be seen in this light, in particular indexing semantic matching. Following Giannakidou (1997) and Giannakidou and Quer (2013), it is reasonable to assume that the NPI contains a lexical polarity feature that marks it as a dependent item sensitive to negation. The following is a definition for NPI after Giannakidou and Quer (2013):

#### (19) *Negative Polarity items*

Denotation:  $[[\text{NPI}(\text{any, kanenas})]] = P(x_d)$ ; where  $x_d$  is a *dependent variable* in need of binding; P stands for the NP predicate.

Licensing: The NPI is an expression whose feature structure contains an uninterpretable POL(arity) feature whose value is *nonveridical*, [*uPol:nonver*]; this feature must enter an Agree relation with a [*iPol:nonver*] head.

In the denotation, the NPI contains a *non-deictic* dependent variable in need of binding.

#### (20) *Non-deictic dependent variable* (Giannakidou, 2011)

A variable  $v$  is non-deictic iff  $v$  cannot be interpreted as a free variable.

We can also think of the dependent variable as a variable that cannot introduce a discourse referent (or, cannot be closed by text level existential closure, as suggested in Giannakidou, 1998). Such a variable won’t be able to get a value from the context, unlike non-dependent variables that can, and will always appear to be “narrow scope”: its distribution will be constrained in contexts where there is an operator it can be bound by. The presence of a dependent variable thus underlies the very essence of NPI-hood.

The presence of a dependent variable creates limited distribution. The dependent variable class includes NPIS—but

<sup>5</sup>NPIS are often discussed in contradistinction to positive polarity items (PPIs). Often a duality is imposed (Progovac, 2005), but Giannakidou (2011, 2017) cautions that the PPI failures are not of the same kind:

- (i) a. John unfortunately died.  
b. #John didn’t unfortunately die.

Metalinguistic denial (Horn, 1989) can rectify PPIs (see Ernst, 2009): *John isn’t here ALREADY, we are still waiting for him*; hence “#” reflects that the PPI is not ungrammatical, but infelicitous in the scope of negation. A failed NPI is ungrammatical, but a failed PPI appears to be only infelicitous, the intended duality between NPIS and PPIs thus does not involve necessarily the same grammatical mechanism. Kriřka (1995) suggests a Gricean explanation that since under negation the NPI is the default, using another form implicates that the alternative scope is intended. Experimental evidence can be useful in distinguishing between NPIS and PPIs, but this will be a topic for future research.

also non-polarity variables such as reflexive pronouns, traces, distributivity markers (reduplicated indefinites in Hungarian; Farkas, 1998), the temporal variable of the subjunctive mood (“temporal” polarity in Giannakidou, 2009), and as recently argued in Grano (2011), subjects of exhaustive control verbs such as *try*, *manage*, etc. This framework imposes an isomorphism between semantics (dependent variable that cannot remain free) and morphosyntax (a dependent variable being a distinct syntactic object from a non-dependent variable). Being a distinct syntactic object means, by *licensing*, that the NPI has a polarity feature. This polarity feature POL is subject to a matching requirement (*Agree*) with the licenser. In other words, the non-deictic variable dependency is lexically encoded in the POL feature. Polarity features have been implied for NPIs since the early days (e.g., Klima’s +affective feature, Giannakidou, 1997 sensitivity features, Zeijlstra’s NEG feature; Chierchia’s + $\sigma$  feature is within the same spirit)<sup>6</sup>.

With NPIs that are more narrowly sensitive to negation, such as e.g., n-words, or the Dutch NPI *hoeven* “need,” it is reasonable to assume that the abstract lexical semantic feature is not POL but [+Neg] (see Lin, 2015 for some recent discussion on the acquisition of this feature and its contrast with acquisition of broader NPIs such as any). In the literature on n-words it is very common to assume [+NEG] (Zeijlstra, 2004; Giannakidou and Zeijlstra, 2017 for an overview). In any case, POL and NEG would be the lexical indexing of the NPI-dependency in the grammatical representation of the NPI word. During the incremental comprehension of a sentence, if NEG or POL are compositionally derived prior to encountering the NPI (i.e., in the licenser), that should lead to a reduced N400 on the NPI. This hypothesis relies on the fact that the amplitude of the N400 evoked by an incoming word indexes the degree to which that word’s semantic features match semantic features that have been preactivated by its context at the time of encounter (Lau et al., 2008; Kutas and Federmeier, 2011; Kuperberg, 2013). The term “pre-activation” has often been associated with active prediction of specific lexical items. (Xiang et al., 2016) use it in a more neutral sense to refer to the activation of relevant semantic features, regardless of whether active prediction or expectation of the upcoming word is at work, ahead of encountering the full linguistic input. In the context of NPI licensing, it is reasonable to assume that, during the incremental comprehension of a sentence, if a semantic [+NEG] feature or [+POL] feature is compositionally derived prior to encountering the NPI, that should lead to a reduced N400 on the NPI word.

In a series of studies (Saddy et al., 2004; Drenhaus et al., 2005, 2006, 2007), a reduced N400 with a central maximum was found on the German NPI *jemals* (“ever”) when it was licensed by negation, compared to the ungrammatical counterpart when *jemals* was not licensed. Similar N400 effects were also found

<sup>6</sup>The concept of dependent variable itself does not necessitate the POL feature. A failed dependent variable can remain a pure presupposition failure. E.g., the presupposition of *#He left* is not satisfied in a female-only domain, but this does not cause ungrammaticality. Dependent variable phenomena are not *all* NPI phenomena (e.g., distributivity markers or controlled subjects are not NPIs); and some polarity phenomena might be purely presupposition failures, as Giannakidou accepts in numerous places (e.g., PPIs, as suggested earlier).

for Dutch (Yurchenko et al., 2013) and English NPIs (Shao and Neville, 1998), a MEG study by Tesan et al. (2012), Xiang et al. (2009), and Xiang et al. (2016). The studies differed considerably in their aims, materials and experimental designs, behavioral task, and test language; but the N400 finding is in line with the idea that NPI licensing involves a semantically driven dependency and suggests that the semantic requirement of the NPIs involves some kind of lexical/morphological feature matching.

Interestingly, another study by Steinhauer et al. (2010) did not find an N400 difference between licensed and unlicensed *ever*<sup>7</sup>; but a crucial difference between their study and the others mentioned above is that Steinhauer et al. had a larger set of licensors in their stimuli, including various negative licensors such as *not*, *without*, *rarely/hardly*, and also licensors that are not negative *per se*, but non-veridical, such as *every*, *before*, *whether*, and *yes-no* questions. It is possible that negative features are only present with classically negative expressions (*no*, *no one*, *not*), and that the varying degree of negativity (or no negativity at all, e.g., with a non-veridical non-negative licenser) causes the N400 effect to be reduced. In Steinhauer et al. the effect could have been watered down by using both negative and non-negative licensors. This, by itself, of course raises the question of how to trace the judgment in the case of non-negative licensors, and this is something that needs to be studied. Overall, what we want to say here is that reduced N400 can be plausibly viewed as a correlate of the underlying lexical semantic matching dependency between the NPI and the licenser that produces an expectancy in the sense of Lau et al. (2008). The range of data confirming this is solid enough to be able to render the N400 effect a predictor. The weakened N400 with non-negative licensors observed in Steinhauer et al. may be a reflection of a matching between an NPI a non-negative licenser, either, as we said, because the non-negative licensors lack [NEG], or because it might have only [+POL]. If the NPI is [+NEG], but the licenser is only [+POL], this is a weaker match. In other words, there appears to be a hierarchy of strength of these lexical features.

At the same time, the presence of P600 effect with NPIs reflects syntactic integration, and the consistent presence of P600 allows us to treat these two components as tied to syntactico-semantic processing.

## Integration Correlates and Two Modes of Licensing

The majority of the NPI studies mentioned above also reported a posteriorly distributed P600 late positivity effect, which was

<sup>7</sup>Multiple NPIs were tested in Steinhauer et al. (2010). N400 difference was only observed for *at all*, but not for *ever* or *any*. Let us also mention that the NPI studies did not always examine NPI licensing alone. To give some examples, Saddy et al. (2004) and Yurchenko et al. (2013) compared PPI and NPI licensing contexts within the same experiment, Xiang et al. (2009) compared NPI licensing with reflexives and some of the German and English studies included an inaccessible licenser condition besides the no-licenser vs. licenser distinction in their manipulation. All these things matter for the possible elicitation of different ERP components.

larger for unlicensed NPIs than for licensed ones<sup>8</sup>. The P600 effect is typically associated with syntactic processing and syntactic complexity, and is reliably elicited by syntactic errors (Osterhout and Holcomb, 1992; Hagoort et al., 1993; see also Osterhout et al., 1994; Kaan et al., 2000; Phillips et al., 2005), or grammatical but syntactically complex constructions (Osterhout et al., 1994; Kaan et al., 2000; Phillips et al., 2005; Gouvêa et al., 2010). Although the precise functional interpretation of the P600 is yet to be determined, a broad generalization that has emerged is that it reflects costs associated with a processing stage in which information from different sources is integrated into one coherent representation. We will therefore take P600 to be an indicator of *integration* cost (Friederici and Weissenborn, 2007; Kuperberg, 2007; Bornkessel-Schlesewsky and Schlewsky, 2008; Van Petten and Luka, 2012). Increased P600 amplitudes signal the detection of an integration error or difficulty<sup>9</sup>.

In the particular context of NPI licensing, multiple streams of information—syntactic, semantic, and in some cases pragmatic, as we mention next—are recruited to construct a grammatical representation that can license NPIs. In an ungrammatical sentence that does not license NPIs, the comprehension system fails to integrate an NPI into the current grammatical representation, and therefore produces a large P600<sup>10</sup>.

Crucially, sometimes NPIs are sanctioned pragmatically, by *implicit* negation. Giannakidou (2006) calls this phenomenon *rescuing* and it is very clear with emotive verbs:

- (21) a. John was surprised that Mary has any friends.  
 b. We are lucky that we got any tickets!  
 c. John regrets that he talked to anybody.

In these sentences, *any* appears without a negative or nonveridical licenser; the emotive verb (*be surprised*, *be lucky*, *regret*), arguably, has a positive thus veridical presupposition that the complement is true. These are, at best, cases of mixed veridicality (Giannakidou and Mari, 2018), but some NPIs, as we see, appear nevertheless (see Baker, 1970; Linebarger, 1987 for earlier data). Giannakidou shows, on the other hand, that the corresponding Greek NPIs are ungrammatical:

- (22) \*Metaniosa pou ipa tipota.  
 regret.1sg that said.1sg anything  
 I regret that I said anything.

As mixed licensers, emotive verbs are highly variable, as confirmed by a recent paper (Duffley and Larrivée, 2015) where it is shown that the appearance of *any* is indeed limited with emotives. Giannakidou argues that with emotives we don't have licensing proper but *rescuing* of the NPI by accessing implicit, i.e., not asserted, negation.

- (23) *Rescuing by NEGATION* (Giannakidou, 2006).  
 A PI  $\alpha$  can be rescued in sentence S, if the global context C of S makes a negative proposition S' available, and (b)  $\alpha$  is in the scope of negation in S'.

Rescuing is proposed by Giannakidou a *secondary* mode of licensing that relies on pragmatic inferencing from the global context, which includes the presuppositions and implicatures of the sentence. Horn (2002) further structures global pragmatics with *assertoric inertia*: one component becomes assertorically inert, and another becomes salient. If the salient component contains negation, NPIs will be licensed. Emotive verbs give rise to a negative inference that has been characterized as an implicature (Linebarger, 1987), or presupposition (Baker, 1970; Giannakidou, 2006, 2016). In the case of emotives, then, the negative meaning arises not from a logical property of the emotive verb—which would render a NEG or POL feature possible—but from implicit negation (*regret* implicates or presupposes that *I wish I didn't*). With rescuing, the NPI needs to access the pragmatic level of representation.

In agreement with the rescuing idea, processing literature treats the appearance of NPIs with emotives as non-canonical, and uses labels such as *illusory* effect (Xiang et al., 2009, 2013), and *erroneous pragmatic licensing* (Yanilmaz and Drury, 2018). Giannakidou (1998) calls rescuing *indirect* licensing; Xiang et al. (2016) further study how licensing proper differs from non-canonical licensing in online computation. They look at the P600 effect. If the P600 indexes the integration effort with which an NPI is licensed, it provides a useful tool to examine whether or not negation in the pragmatics is treated by the comprehension system as an equally viable licenser.

Combining observations from the N400 and the P600 time windows, Xiang et al. state that they can construct a complete picture as to when and how negation is computed and used for grammatical purposes. The N400 reveals information about whether a negative meaning is established incrementally in context; the P600 assesses whether negative meaning, if available, can be immediately adopted to serve the grammatical function of NPI licensing.

Xiang et al. in their experiment 1 found the following. All (*no*, *few*, *only*) conditions that contain a legitimate licenser showed, expectedly, a qualitatively similar N400 reduction during the 300–400 ms time window at the critical NPI, relative to the unlicensed condition. They also by and large showed a reduced anterior negativity compared to the unlicensed condition during the late 700–900 ms time window. However, during the P600 time window, although conditions under *no*, *few*, and *only* showed

<sup>8</sup>Only two studies did not find a P600: Saddy et al. (2004) and Yurchenko et al. (2013). The original data from Saddy et al. were reanalyzed in Drenhaus et al. (2006) using a symbolic resonance analysis, and a hidden P600 was discovered. Yurchenko et al. acknowledged that the lack of a P600 may be due to insufficient power in the data, as well as, potentially, to task-specific effects.

<sup>9</sup>Importantly, we do not imply seriality in processing stages as e.g., in the three-phase model of Friederici (1995, 2002), Friederici and Kotz (2003), Friederici and Weissenborn (2007). Steinhauer and Drury offer a critique of that model, concluding that “the three-stage architecture of Friederici’s model may have to be modified as there seems to be little evidence for a first phase exclusively dedicated to phrase structure processing. Moreover, context-driven top-down processing may play a larger role than assumed by the current version of this model.” (Steinhauer and Drury, 2012, p. 154). Our point in this paper is simply to raise awareness that N400 and P600 are physical correlates of licensing; we remain agnostic wrt to architecture that clearly requires more study.

<sup>10</sup>It is worth mentioning here the so-called “semantic P600,” which sometimes appears with an absence of N400 effect. Chow and Phillips (2013) offer a detailed discussion arguing that semantic P600 is compatible with the long-held assumption that online semantic composition is dependent on surface syntax.

qualitatively similar patterns involving a smaller P600 amplitude relative to the unlicensed condition, the emotive predicate condition yielded a P600 as large as the P600 in the unlicensed condition. This result sets rescuing with emotive predicates apart from licensing proper, and supports the thesis that rescuing does not involve syntax-semantic integration (as that would predict no P600). Trying to access negation in the pragmatics produces the effect. Giannakidou (1998, 2006) is the only currently available theory that predicts this fact, as it is the one that posits two qualitative distinct modes of licensing (licensing proper which involves syntax-semantics integration, vs. rescuing which involves pragmatics).

Let us note that Xiang et al. (2016) consider the possibility that the P600 effect may be due to fact that with emotives the NPI is embedded in a different clause, unlike in the other conditions where licenser and NPI are in the same clause (*No student said anything* vs. *Maria regrets that she said anything*). This is also a question raised by a reviewer. Xiang et al. therefore designed a second, self-paced reading experiment aiming to examine whether the ERP pattern could be replicated in a different paradigm, and also to assesses whether the additional processing cost found on the emotive condition is due to its NPI licensing properties or to the other possible sources of processing complexity (such as e.g., embedding). They conducted a  $2 \times 5$ -design self-paced reading study, the results of which rule out the possibility that the observed effects among the NPI conditions should be attributed to independent structural or contextual differences among different conditions. Of course, caution needs to be taken in drawing parallel relations between ERP and self-paced reading results; but the fact that the same costs, with the same relative timing, are observed in the NPI conditions from both the ERP data and the reading time data suggests that they are comparable measures to examine the online processing of NPI comprehension (see Xiang et al., 2016 for the experimental details).

What are the overall conclusions from this discussion? We suggest the following:

1. Experimental, specifically ERP, methodologies are useful in establishing physical, quantitative correlates of interface intuitions that can serve as criteria for distinguishing aspects of the linguistic judgment.
2. NPIs exhibit a complex judgment that involves integrations from multiple levels: syntax, semantics, and potentially pragmatics, as is in the case of rescuing.
3. Reduced N400 can be understood as the physical correlate of semantic licensing, i.e., as a matching relation between the NPI and its licenser.
4. P600 can be understood as the physical correlate of the syntactic aspect of licensing, i.e., integration.
5. The contrast between rescuing and licensing is real, and observed at the expected level of P600 as an integration effect.

Let us remind again that the N400 is not an index of felicity (as one would think, for instance by reading only, Shao and Neville, 1998), but of predictability, expectancy and matching. Overall, the ERP methodology allowed us to disentangle the key aspects

of grammatical judgment with NPI licensing, in ways that would have been impossible with mere intuition. This, we believe, is a promising result—sufficient in itself to generate and enhance interest in pursuing these methodologies further for such types of phenomena.

Before we move on to different kinds of experiments, we want to offer a few comments on why, we think, ERP methodology is useful for NPI licensing, or, in other words, why tracking processing in time matters for this type of phenomenon. As it became clear, NPI licensing is a long-distance dependency: it requires ability of the processor to integrate material that may not be locally adjacent (*No student saw anything*). Other well-known long distance dependencies are wh-dependencies (*Who did Ariadne see t?*), and the antecedent—anaphor relations (*Ariadne likes talking to herself*) that we talked about at the beginning. In all these cases, e.g., upon encountering the NPI, the trace, or the anaphor, the processor must assess the structure that has already been processed. This raises the question of how structured representations are encoded in memory, and how representations are retrieved to extract information. Hierarchically structured representations must be tracked during language processing in order for the parser to “accurately single out grammatically licit antecedents, and representations of structure in memory must be organized in such a way that retrieval operations can make appropriate decisions about acceptable or unacceptable targets” (Xiang et al., 2009, p. 40). This entails that observing processing in time is an effective technique for assessing long-distance phenomena—and although the three phenomena mentioned here are distinct, they all involve explicit reference to previously processed lexical items and structure, hence they benefit from tracking processing on time. ERP methodology can thus provide a secure take, we believe, on this type of syntax-semantics integration.

We move on now to different methodology. We show that mere acceptability judgment tasks can also be useful in revealing more sharpened intuitions with NPIs.

## EMPIRICAL VARIATION AND THE SCALE OF NEGATIVITY

NPIs, as mentioned earlier, are known to be licensed by classical and minimal negation, as well as nonveridical expressions that are not negative. NPI licensers can thus be viewed as being of variable strength when it comes to negativity, an idea expressed for the first time in Zwarts (1996). Recall the NPI licensers in **Figure 1**. Negation itself is the strongest licenser, but minimal negations (merely downward entailing such as *few*) are weaker, and non-negative licensers (questions, modals, etc.) are the weakest, with zero negativity. In the Zwarts and Giannakidou framework (**Figure 1**), negativity emerges as a gradient property, i.e., a scale:

- (24) Scale of Negativity (Zwarts, 1996; Giannakidou, 1997)  
 <non-negative, mere downward entailment, antiadditive, antimorphic>

*Nobody* is more negative than *few* (it satisfies three de Morgan laws, but *few* only satisfies two). And sentence negation is antimorphic, the strongest negation satisfying all de Morgan relations. Non-negative elements have zero negativity, which means that none of the negative laws apply. Chatzikontantinou et al. (2015) set out to test the predictions of this theoretical proposal. They also included the emotive verbs we discussed earlier and *only* which are of mixed veridicality.

We give below a brief description of their task. Seventy five native speakers of Greek in Greece were presented with 30 statements-pairs and were asked to point on a 1–5 scale whether the second statement is an acceptable continuation of the first. The second statement contained the NPI *pote* “ever”. The participants were asked to judge if the sentence is acceptable. Materials included five types of (S2) continuations differing on the (non)licensors. Sentence structure was kept as similar as possible e.g.,:

- (25) (S1) Special effects are expensive.  
(S2) {Elaxisti/ It surprised me that/Only} skinothetes  
{Very few/It surprised me that/Only}directors  
xrisimopiisan *pote* idhika efe.  
used ever special effects.

Negative quantifiers are n-words in Greek that must co-occur with negation creating negative concord (mentioned earlier). This was the condition used for negation:

- (26) Kanenas skinothetis dhen xrisimopiise *pote* idhika efe.  
no director not used ever special effects  
No director has ever used special effects.

It was expected that licensing proper would be the most solid judgment—and that, if Giannakidou and Zwarts’ view of negativity is correct, we would have some variation in the data even with licensing. Overall the strength of licensors is:

- (27) Negative strength of licensors: “>” indicates “stronger than”  
negation > very few > only > factives > no-licenser

The results are the following:

Licenser	Dependency	Acceptability Rate
sentential negation	Licensing	4.7
Ελάχιστοι ‘very few’	Licensing, DE	3.5
μόνο ‘only’	Rescuing	2.5
emotive	Rescuing	2.3
no-licenser	–	0.3

These data show that each licenser was associated with different degree of acceptability, with all *t*-tests comparing conditions being highly significant. Chatzikonstantinou et al. ran a 1 way Anova and the results showed a main licenser effect ( $F = 121.337$ ,  $p < 0001$ ) which suggests that it matters what licenser you are. The analysis showed that apart from the comparison between *emotive factives* and *only* ( $p = 0.14$ ) all other comparisons are statistically significant ( $p < 0.0001$ ) which suggests a division among the NPI licensors.

These findings confirm the scalar negativity hypothesis. They indicate a distinction between licensing by negation and licensing by downward entailment (*elaxistoi* “very few”), and a difference between licensing proper [both cases in (a)] and the rescuing we discussed earlier. These differences cannot be captured in accounts that do not differentiate between modes of (licensing vs. rescuing; e.g., von Stechow, 1999) and strength of licensors. And, importantly, the variation in the data, again, could not have been revealed without the judgment task.

Hence, for interface phenomena such as NPI licensing, even simple quantitative methods can be helpful in revealing the empirical patterns that are relevant for theorizing. At the same time, as shown earlier, ERP methodology enables establishing physical, quantitative correlates of interface intuitions that can serve as criteria for distinguishing aspects of the linguistic judgment. NPIs exhibit a complex judgment that involves integrations from multiple levels: syntax, semantics, and potentially pragmatics, as is in the case of rescuing. Reduced N400 can be understood, we argued, as the physical correlate of the semantic aspect of licensing, i.e., as a matching relation between the NPI and its licenser, while P600 can be taken to index the syntactic aspect of licensing and the cost of integration.

## INTONATION AND MEANING: DISAMBIGUATING SCALAR COMPONENTS

We now want to study another kind of interface judgment: the one derived from prosody and semantics interface. Interactions between prosody and scope have been consistently noted in the literature on meaning and intonation, since they were first addressed in Jackendoff (1972). Actually, the question why and under which circumstances scope inversion is possible has provoked a fair amount of approaches, see references in Horn (1989: 226ff). Jackendoff (1972) noted that the example in (28), in an out-of-the-blue context is ambiguous between the interpretation in (28a) and (28b), depending on the scopal relation of the universal quantifier *all* and the sentential negation.

- (28) All the men didn’t go  
a.  $\forall > \neg$ : no man went  
b.  $\neg > \forall$ : some men went

As soon as intonation changes, this affects the sentence and only one of the readings is available. If the sentence is uttered with the rising contour, expressed by the lines below the example in (29a), the sentence is interpreted with negation taking scope over the universal quantifier *all*, i.e., *some men went*, while if the sentence ends with a falling pitch contour as in (29b), it is the universal quantifier that takes scope over the sentential negation and means that *no men went*. Büring (1997) and Krifka (1998) account for this data by making use of contrastive topicalization, that according to them, involves scope inversion in these cases.

We will not get into discussing these proposals here, as we just want to show the facts.

- (29) a. ALL the men didn't go. [B accent:  $\neg > \forall$ ]  
 b. ALL the men didn't go. [A accent:  $\forall > \neg$ ]

Other authors have also worked on the topic of the interaction of quantifier scope and prosody, e.g., Martí (2001), Kennelly (2003; 2004), Etxeberria and Irurtzun (2004), and Jackson (2007), etc.

In some other cases, prosody and information structure eliminate the ambiguity of a quantificational element. Thus, so-called weak Qs are assumed to be ambiguous between a cardinal and a proportional reading (cf. Milsark, 1977). Thus, a sentence like (30) exemplifies the two possible interpretations of *some* (cf. Partee, 1988).

- (30) Some girls are playing basketball.

On its proportional interpretation, the meaning of *some* can be paraphrased as *some, but not others*, and it is synonymous with the partitive *some of the children*. This interpretation is felicitous only when the set of students is already under discussion. On its cardinal reading on the other hand, *ume batzuk* is not paraphrasable as a partitive, and the weak Q just makes reference to a quantity.

What appears to be extremely interesting is that as soon as prosody affects the sentence and the weak Q *some* is focalized, as exemplified in (31), the cardinal interpretation disappears and *some girls* can only be interpreted proportionally as *some, but not others*.

- (31) [Some]<sub>F</sub> girls are playing basketball.

Here again, as was the case in ambiguous sentences with two Qs, the focalization of the weak Q disambiguates the sentence.

We can find a similar effect with numerical noun phrases like three *beers*. The sentence in (32) is ambiguous between an *at least* interpretation and an *exactly* interpretation.

- (32) I drank three beers.

Horn (1972) analyses the *at least* interpretation as an implicature, as shown by the felicity of (33) where the continuation eliminates the exactly interpretation of *three beers*.

- (33) I drank three beers, in fact, I drank seven.

Crucially, as soon as the prosody of the sentence is changed by focalizing the numerical expression, we only get the *exactly* interpretation as shown by the ungrammaticality of (34).

- (34) \*I drank [three]<sub>F</sub> beers, in fact, I drank seven.

We get a similar effect here too, that is, as soon as we focalize the numerical expression one of the possible interpretations (the one we obtain via implicature, according to Horn, 1972) disappears.

The effect played by prosody has also been studied in other contexts such as answers to polarity questions (see Li et al., 2016, etc.), factive presupposition projections (Beaver, 2010;

Tonhauser, 2016; Simons et al., 2017, etc.), etc. In this section we concentrate on scalar meanings that can be created by making use of prosodic marking. We first present production data from Chatzikonstantinou (2016) showing distinct prosodic properties correlating with scalar readings of Greek NPIs (5.1); then, we discuss a production and perception study by Etxeberria and Irurtzun (2015) on the prosodic disambiguation of the scalar/non-scalar readings of sentences containing the additive particle “ere” in Basque. In both cases we see prosody influencing or affecting semantic interpretation, and experimental methodology is, again, instrumental in refining and offering physical correlates of linguistic intuition. We will not discuss what explains the facts best, we will only concentrate on describing the facts.

## Greek NPIs

It has been a common observation that Greek exhibits a difference between two variants of NPIs distinguished by “emphatic accent” (Veloudis, 1982; Giannakidou, 1997 et seq., Tsimpli and Roussou, 1996); upper-case indicates the obligatory presence of prosodic prominence in a phrasal context. The emphatic NPI is interpreted as an n-word and participates in negative concord (Giannakidou, 1998, 2000), i.e., it requires negation and cannot appear in questions, unlike non-emphatic NPIs:

- (35) a. kanenas/KANENAS “anyone, anybody/  
no-one, nobody”  
 b. tipota/TIPOTA “anything/nothing”  
 c. pote/POTE “ever/never”  
 d. puthena/PUTHENA “anywhere/nowhere”  
 e. katholu/KATHOLU “at all/not at all”
- (36) a. Dhen idhe kanenan o Janis.  
not saw NPI.person the John  
“John didn't see anybody.”  
 b. \*Idhe kanenan/KANENAN o Janis.  
saw.3sg NPI.person the John  
 c. Dhen idhe KANENAN o Janis.  
not saw NPI.person the John  
“John didn't see *anybody at all*.”  
 d. Idhes kanenan/\*KANENAN?  
Saw.2sg NPI.person  
Did you see anybody?  
 e. An dhis kanenan/\*KANENAN  
if see.2sg NPI.person  
If you see anybody

Giannakidou and Yoon (2016) present a number of arguments showing that emphatic NPI gives rise to intensified, *scalar* negation, akin to *anybody at all* (38c), while the non-emphatic NPI (previously encountered) is non-scalar. To understand the contrast, consider the following scenario (from Giannakidou and Yoon, 2016):

- (37) Context: Maria is supposed to read some articles this week for *Semantics 2*, of which only one is required (the others are optional). Maria is notoriously late in doing



her readings, usually doing the minimum. Her friend Ariadne asks the day before class:

- Ariadne: Dhiavases toulastiston to ypoxreotiko arthro?  
 “Did you read at least the required article?”
- Maria: a. Ax, oxī! Dhen dhiavasa KANENA arthro!  
 Ah, no! not read.1sg NPI.det article  
 b. Ax, oxī! #Dhen dhiavasa kanena arthro!  
 “I didn’t read any article at all!”

The non-emphatic NPI in (37a), in contrast to the emphatic one in (37b), is infelicitous. By using the *at-least* phrase in the question, the question forces a scalar reading (the required article is the most likely to read, or the least likely to ignore). The non-emphatic NPI is an odd device in the scalar context, while the emphatic NPI is fine. It is useful to see the parallel with *any*: *any* with devices such as *at all* differs from bare *any*, which can be used in statements that are rather neutral (see Duffley and Larrivé, 2012 for recent discussion).

(38) If you find any typos in this text, please let us know.

(39) Hitting any key will reactivate the screen.

Duffley and Larrivé (2012, p. 30) conclude that: “a good number of common uses of *any* are not amenable to a scalar interpretation at all,” as we can see in the examples above. In Greek NPIs, the *at all* intensification happens purely prosodically.

The prosodic distinction remained a theoretical generalization for many years since Veloudis’ and Giannakidou’s initial observations—but Chatzikonstantinou (2016) conducted production and comprehension experiments suggesting that the difference is observed empirically. For the purpose of this paper, we consider one of his production experiments. 30 native speakers (15 male, 15 female) of Greek were recruited aged from 22 to 55 (mean age 32). All of them were born in Greece and had completed at least the 12 obligatory years of school education while some of them had a higher education degree. The task was administered individually to each participant on a computer screen, and included slides with the scalar (such as

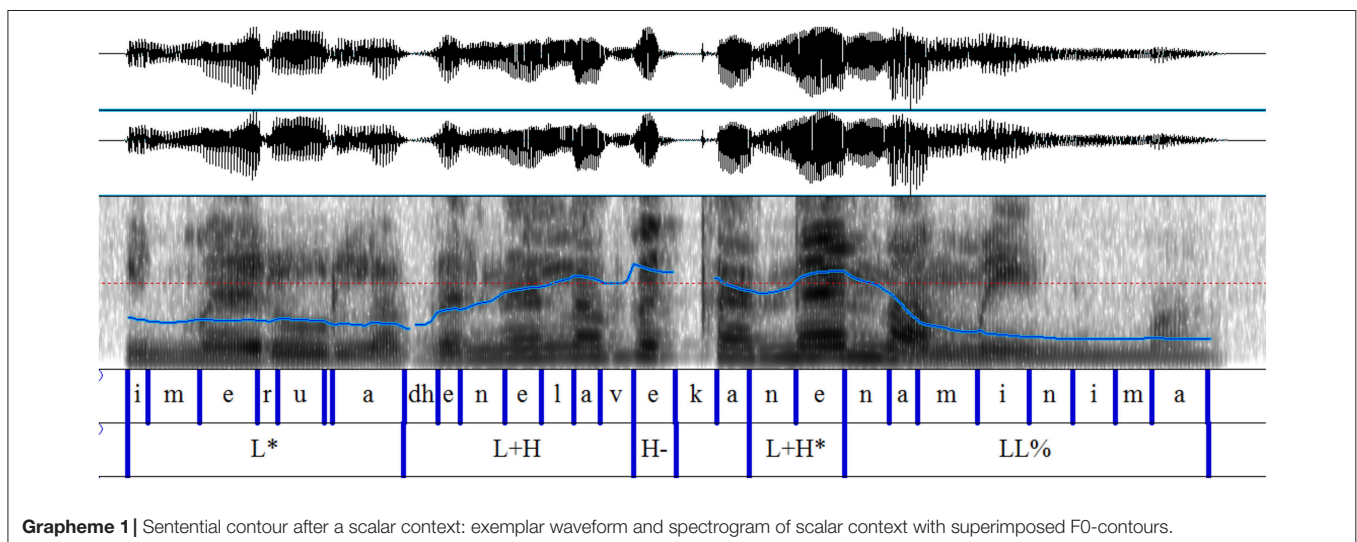
(37) above) or non-scalar contexts. Each slide consisted of the context and a target sentence in bold font. The sentences were written in the Greek alphabet, and the whole information fitted in one slide. Each participant was instructed to first read the context, try to get a good understanding of it. There was no time limitation for this and participants were told that they can read the text as many times as they want. Upon this, the instructions guided them to read the target sentence aloud as if it was a kind of summary or continuation of what has been narrated in the context. Participants were also informed that during the whole process that their voice would be audio recorded. In each session, the experimenter was present and assisted with the procedure.

The scalar and non-scalar contexts presented the images we see in **Graphemes 1, 2** (from Chatzikonstantinou, 2016). The contours are different, the most notable difference being the Low plateau within which the non-scalar NPI is realized. There is a High peak in the beginning of the utterance and then a deaccentuation till the end. The alignment of the High peak—here aligned with the S in the beginning of the utterance—varied as often it was aligned with the negative marker bearing a more typical negative contour. No intermediate phrase was observed as no pauses during the utterance were perceivable.

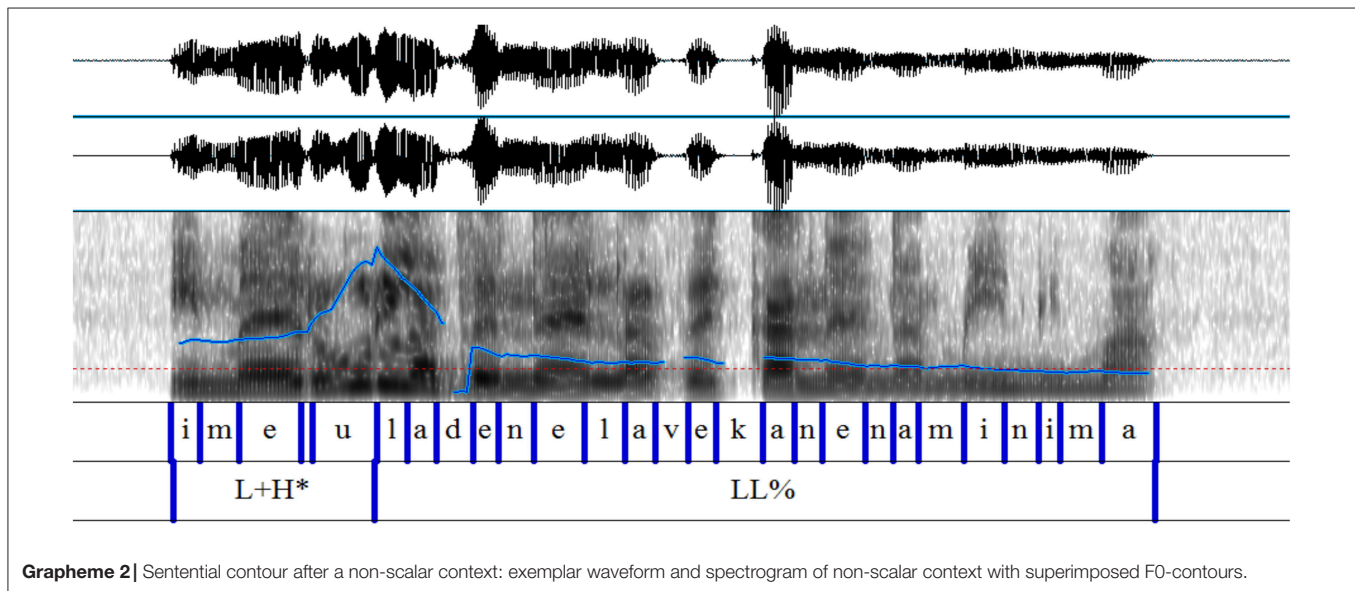
Sentential contours are distinct in the two paradigms. The pitch contour (F0) looks quite different: the emphatic is associated with a L+H\* (the H\* is aligned with the stressed syllable) and then a fall—but the non-emphatic has a flat intonation (and also the part before and after it).

Chatzikonstantinou investigated further the F0 and ran a two way Anova (Scalarity (scalar, non-scalar) x Tonal Target (/e/ and /n/) contained in kanenas [the bolded characters are the *tonal targets*]). There was a significant main effect of *scalarity* on the pitch value produced,  $F_{(1, 83)} = 104,097$ ,  $p < 0.001$ . There was also a main effect of *tonal target*  $F_{(1, 83)} = 18,859$ ,  $p < 0.001$  and an interaction effect between *scalarity* and *tonal target*  $F = 17.917$ ,  $p < 0.001$ . The result suggests that Pitch is a robust acoustic cue that differentiates between a scalar and a non-scalar NPI.

Finally, duration measures were taken from /e/ and /n/ again and ran a two way Anova. The results show that there was a



**Grapheme 1** | Sentential contour after a scalar context: exemplar waveform and spectrogram of scalar context with superimposed F0-contours.



significant main effect of *scalarity* on the duration  $F_{(1, 83)} = 51,283$ ,  $p < 0.001$  which suggests that it makes a difference whether you are a scalar or a non-scalar NPI. A marginally significant effect on tonal target was also found [ $F_{(1, 83)} = 3,964$ ,  $p < 0.05$ ]. There was no interaction between the two factors ( $F = 1,621$ ,  $p = 207$ ).

To sum up, NPIs can be scalar and non-scalar, and the difference surfaces in prosodic properties; for more extensive discussion see Chatzikonstantinou's thesis, chapter 3. The important conclusion here is that the theoretical postulate of two prosodic profiles for Greek NPIs, which has been a mere theoretical statement for about 30 years, is actually confirmed by experimental data. This carries significant promise as one further explores the syntax-semantics and prosody interaction. We find next a similar pattern about the role of prosody in bringing about scalar and non-scalar focus in Basque.

## The Basque Additive Particle

In Basque, the ordinary additive particle, *ere*, is used to express both a simple additive, non-scalar value (akin to English *too/also*) and a scalar additive value (akin to English *even*). In fact, *ere* is the only particle available in Basque to produce either simple additives or scalar additives (as opposed to other languages that have different items in the lexicon for different readings, e.g., *too/also* and *even* in English, *también* and *incluso* in Spanish, *aussi* and *même* in French, etc.). Thus, in Basque, a string like (40), with the same lexical items and the same word order can obtain either a simple additive reading or a scalar additive reading:

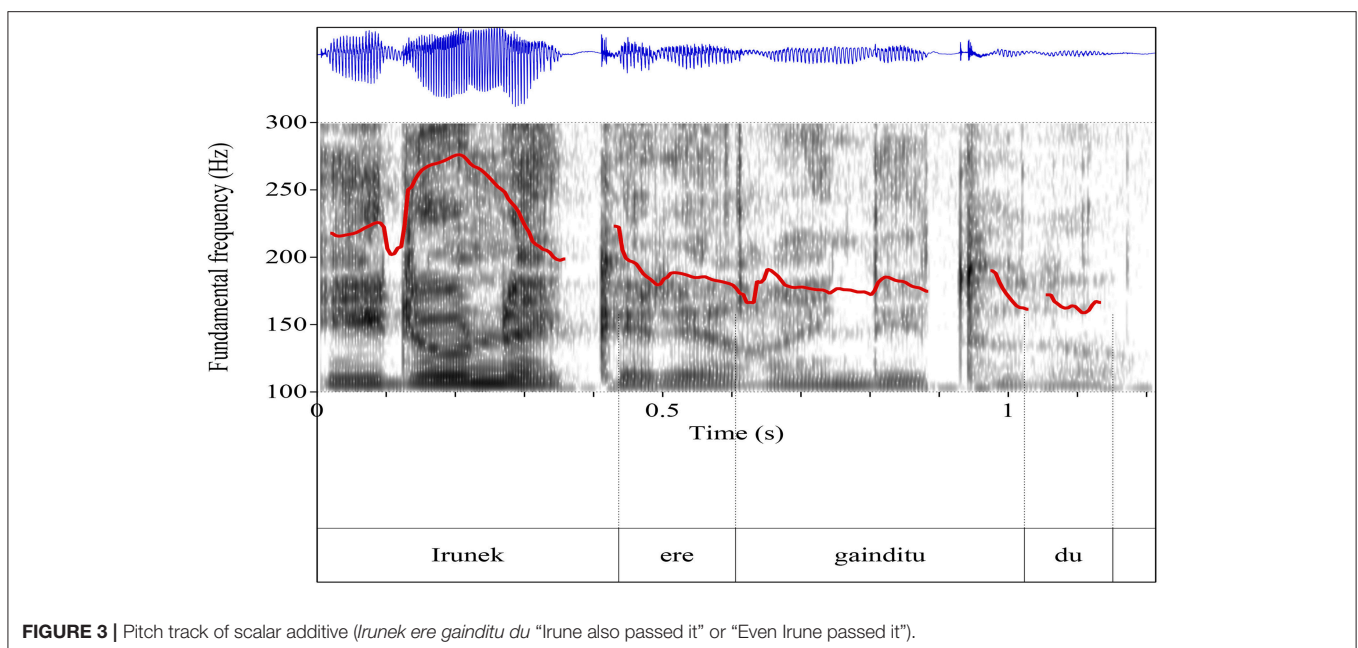
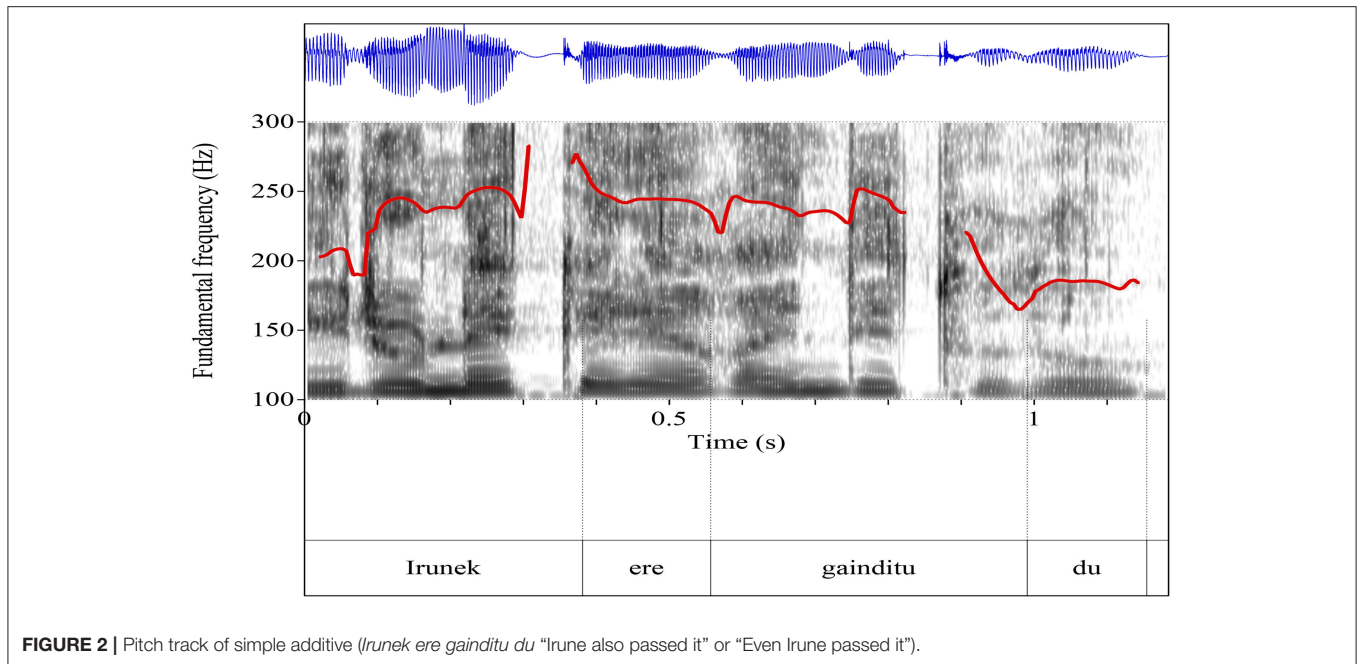
- (40) Mikel ere joan da.  
 Mikel ere go aux  
*Simple*: Mikel left too.  
*Scalar*: Even Mikel left.

At a first look, it would seem then that sentences containing the particle *ere* are completely ambiguous between the simple and the scalar additive interpretations. However, Etxeberria and Irurtzun (2015) show that prosody (the

placement of the nuclear stress) is the key factor for teasing apart the two readings in (40). In order to verify the effect that prosody plays in disambiguating the simple and the scalar additive readings of *ere*, Etxeberria and Irurtzun designed two experiments: (i) a production experiment which aimed to test the prosodic patterns associated to each of the readings and, (ii) a perception experiment, a sentence-comprehension task where subjects had to judge the possible readings of utterances with the additive particle with varying prosodic patterns.

In the production experiment, they asked native speakers of Basque to utter pairs of identical strings corresponding to simple additive and scalar additive interpretations after presenting them a context (via written text) that clearly favored one of the two possible interpretations. They made use of three different strings, and two conditions per string, "Simple" and "Scalar," and all of them contained the same syllable in the accented positions in the element preceding the particle *ere* (/ru/) and the verb following it (/di/). All participants read the same set of sentences. One of the strings they used is exemplified below between brackets "< >."

- (41) a. Simple (**Figure 2**):  
 Mertxek azterketa gaindittu do, eta <Irunek ere gaindittu do>.  
*English translation*: Mertxe passed the exam, and <Irunek ere (=too) passed the exam>.
- b. Scalar (**Figure 3**):  
 Irune klaseko txarrena da, askokatik gainea. Askotan pasatzen da klaseko danok azterketetan nota ona ateatzea eta beak suspenditzea. Halare, lehenon jarri ziguten azterketa hain erraza izan zan, <Irunek ere gaindittu dola>.  
*English translation*: Irune is, by far, the weakest in our class. Often times, we all get good grades and she gets an F. However, the exam that we got the other day was such an easy one that <Irunek ere (=even) passed the exam>.



They measured syllable duration (in ms.), F0 mean and maxima (in Hertz), and intensity mean and maxima (in dB.) in the three syllables, as well as the F0 declination between F0 maxima in syllables /ru/ and /di/. The measurements show a clear difference between strings uttered in the simple condition and strings uttered in the scalar condition in that the stress associated to the element preceding the particle *ere* in the scalar condition is stronger (in F0 and intensity) than in the simple condition and they argue that this is a signature of their focal nature, since narrow focus is associated to nuclear stress in Basque.

Furthermore, they also show that in the Scalar condition the region following *ere* displays reduced F0 values in comparison to the Simple condition, which they linked to the well attested effect of postfocal pitch compression (cf. Elordieta, 1997, 2003; Elordieta and Irurtzun, 2009; Irurtzun, 2013; Hualde and Elordieta, 2014). They conclude that speakers associate different prosodic patterns to different interpretations of the same string, which is a remarkable fact because despite the fact that the contexts of utterance were unambiguous enough so that speakers would not convey any differences

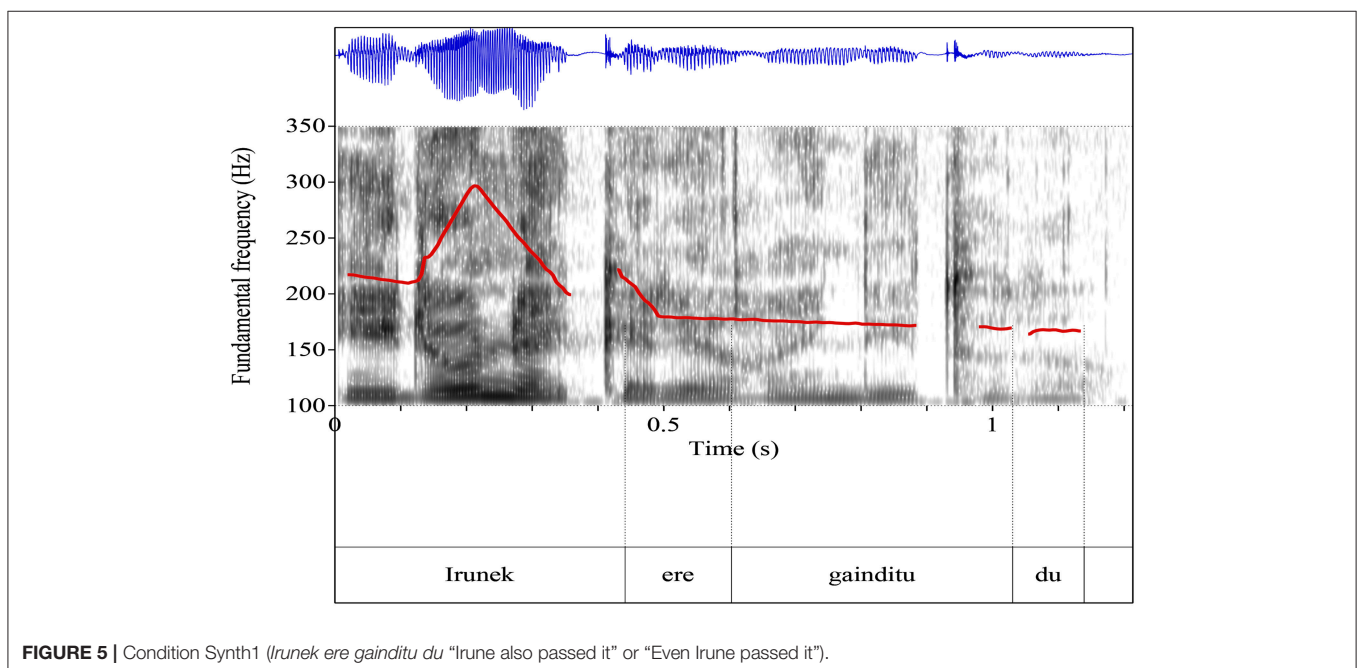
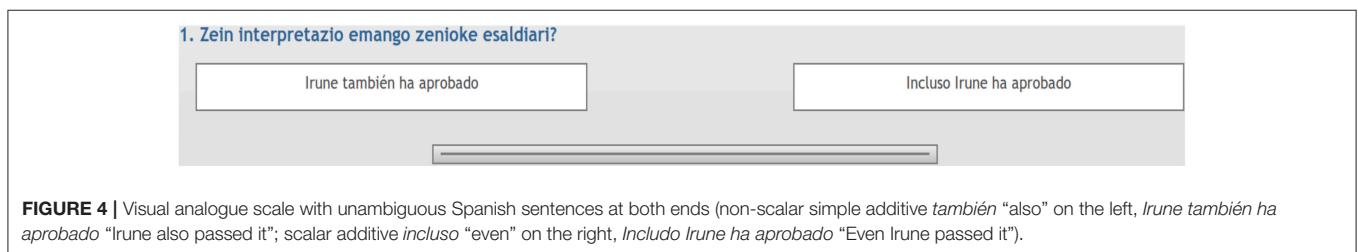
in their prosodic marking, i.e., despite the fact that the exact interpretation of *ere* (scalar and non-scalar) could be inferred from the context alone, speakers produce different tunes.

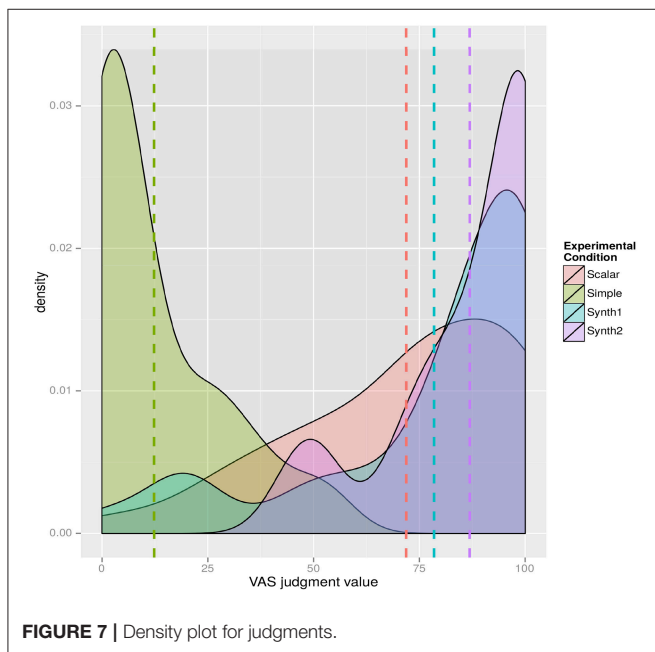
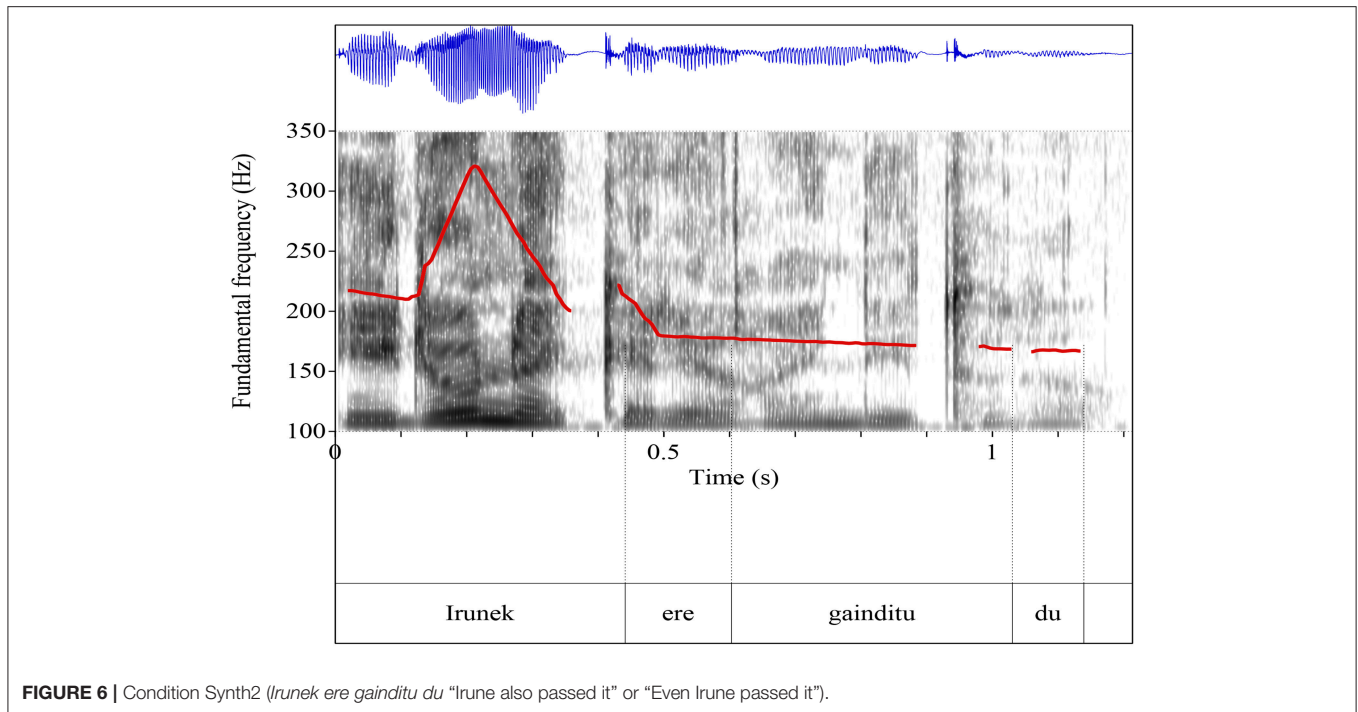
In order to check whether this intonational pattern is enough to convey the intended meaning, they run a perception experiment. For the perception experiment they designed a magnitude-estimation task with the help of a Visual Analogue Scale (VAS) with unambiguous interpretations at both ends (since all Central Basque speakers are bilingual speakers of Spanish and Basque, unambiguous Spanish sentences at both ends were used (with *también* “also”–*Irunek también ha aprobado* “Irunek also passed the exam”– and *incluso* “even”–*Incluso Irunek ha aprobado* “Even Irunek passed the exam”–) (Figure 4).

Participants had to listen to three strings uttered with two different interpretations (simple and additive) in mind, which were taken from the productions of a participant in production experiment. Besides, for the item *Irunek ere gainditu du* “(Even) Irunek (too) passed the exam,” they created an additional pair of test items: Condition Synth1, a manipulation of the item for “Scalar” by stylizing F0, raising the peak of the pitch accent in the subject by 25 Hz, and flattening the postaccentual region

(Figure 5), and Condition Synth2, a manipulation of the item for “Scalar” by stylizing F0, raising the peak of the pitch accent in the subject by 50 Hz and flattening the postaccentual region (Figure 6).

These experimental items (the same string with the same lexical items and the same word order in all cases, i.e., *Irunek ere gainditu du* “Irunek also passed it” or “Even Irunek passed it”) were offered to the participants without any kind of context and participants had to judge the range of possible interpretations of each utterance in the VAS by cutting the judgment line in two: (i) if they thought that the utterance was ambiguous and that it could equally represent the two readings, subjects were instructed to place the delimiter in the middle of the line; (ii) if they thought that it represented more the reading to the left, but still leaving some plausibility to the reading to the right they should place the delimiter on whichever place they felt on the left; (iii) alternatively, if they judged that the utterance was unambiguous in the other direction, they should place the delimiter more to the right. Subjects were explicitly instructed that they could place the delimiter at any point in the line. Etxeberria and Irurtzun controlled the validity of the technique with completely unambiguous fillers that could only be given one interpretation and hence





should be placed at the extreme left or right boundary of the line.

The results of the VAS (from 0 to 100, 0, the value on the leftmost edge, 100 the value on the rightmost edge) show a clearly skewed distribution (Simple  $M = 12.31$ ,  $SD = 15.58$ ; Scalar  $M = 71.88$ ,  $SD = 26.37$ ). The results are very interesting in that the stronger the accent the interpretation gets more scalar (Synth1  $M = 78.47$ ,  $SD = 28.74$  and Synth2  $M = 86.88$ ,  $SD = 17.30$ ).

As **Figure 6** shows, responses to different conditions show a different behavior, with clearly skewed distributions, significantly so in the cases of conditions Simple, Synth1 and Synth2 (**Figure 7**).

Thus, the paper by Etxeberria and Irurtzun shows that constructions with *ere* can vary in their interpretations between the simple and the scalar additivity readings but that these two readings differ depending on where the focal intonation, i.e., the nuclear stress, is placed. As a consequence, the two interpretations that can be obtained in Basque in sentences with *ere* are not to be considered as genuine ambiguity. In other words, there is a clear correspondence between the nonfocal or focal nature of the element preceding the additive particle *ere* and interpreting *ere* the sentence as simple additive or scalar additive. This shows that Basque make use of prosodic properties to disambiguate the scalar or non-scalar interpretations of the additive particle *ere*.

## GENERAL CONCLUSIONS

Our goal in this article was to discuss one of the major questions addressed in this volume, namely if experimental methodologies are helpful in assessing linguistic data and theories about them. We reviewed some recent key literature on the licensing of negative polarity items (NPIs) and on the prosody-semantics interface. We found indeed that experimental methodologies allow us to establish and disentangle patterns and physical correlations of linguistic intuition that would otherwise remain undetected. The phenomena we reviewed involve what we called interface judgment, which is the intuition produced by integrating multiple levels of linguistic representation. We

addressed three main areas of integration involving syntax, semantics and prosody.

ERP methodology, in particular, by tracking processing in time, was useful in establishing physical, quantitative correlates of NPI licensing. Reduced N400, we suggested, can be understood as the physical correlate of semantic licensing, and the observed P600 is an integration effect. In section Empirical Variation and the Scale of Negativity we saw that a mere acceptability judgment task was useful in revealing more sharpened intuitions about degrees of strength of NPI-licensors.

We chose NPIs and the focus particle EVEN because these are areas that we have studied in our previous works, and in the article we synthesized among results that included our own. Overall, the experimental methodologies allowed us to tease apart the key aspects of grammatical judgment with NPI licensing, including prosodic properties of NPIs. In addition, disambiguation of scalar and non-scalar readings of a single word (Greek NPI, Basque *ere*) was clearly established with the aid of phonological experimental observation. Our overall conclusion is that we can be hopeful that experimental methodology can be a helpful tool for interface judgment in revealing the actual empirical patterns that are relevant for theorizing.

## REFERENCES

- Aloni, M. (2007). Free choice, modals and imperatives. *Nat. Lang. Semantics* 15, 65–94. doi: 10.1007/s11050-007-9010-2
- Baker, C. L. (1970). Double negatives. *Linguist. Inq.* 1, 169–186.
- Beaver, D. (2010). “Have you noticed that your belly button lint colour is related to the colour of your clothing?” in *Presuppositions and Discourse: Essays Offered to Hans Kamp*, eds R. Bäuerle, U. Reyle, and T. E. Zimmerman (Bingley: Emerald Group Publishing Limited), 65–100.
- Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Res. Rev.* 59, 55–73. doi: 10.1016/j.brainresrev.2008.05.003
- Büring, D. (1997). The great scope inversion conspiracy. *Linguist. Philos.* 20, 175–194. doi: 10.1023/A:1005397026866
- Chatzikonstantinou, A. (2016). *Semantic and Prosodic Processing of Negative Polarity Items in Greek*. Ph.D. University of Chicago, Chicago, IL.
- Chatzikonstantinou, A., Giannakidou, A., and Manouilidou, C. (2015). “Gradient strength of NPI-licensors in Greek,” in *Paper Presented at the 12th International Conference on Greek Linguistics* (Potsdam: University of Potsdam).
- Chierchia, G. (2006). Broaden your views. Implicatures of domain widening and the ‘logicality’ of language. *Linguist. Inq.* 37, 535–590. doi: 10.1162/ling.2006.37.4.535
- Chierchia, G. (2013). *Logic in Grammar*. Polarity, Free Choice, and Intervention. Oxford: Oxford University Press.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Chomsky, N. (1964). *Current Issues in Linguistic Theory*. The Hague: Mouton.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris Publications.
- Chow, W.-Y., and Phillips, C. (2013). No semantic illusion in the “Semantic P600” phenomenon: ERP evidence from Mandarin Chinese. *Brain Res.* 1506, 76–93. doi: 10.1016/j.brainres.2013.02.016
- Drenhaus, H., Beim Graben, P., Saddy, D., and Frisch, S. (2006). Diagnosis and repair of negative polarity constructions in the light of symbolic resonance analysis. *Brain Lang.* 96, 255–268. doi: 10.1016/j.bandl.2005.05.001
- Drenhaus, H., Blaszczyk, J., and Schutte, J. (2007). “Some psycholinguistic commentson NPI licensing,” in *Proceedings of Sinn und Bedeutung*, Vol. 11, ed E. Puig-Waldmuller (Barcelona: Universitat Pompeu Fabra), 180–193.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

UE would like to acknowledge that this research benefited from the following grants: IT769-13 (Eusko Jaurlaritza-Basque Government), UFI 11/14 HiTeDI (UPV/EHU), EC FP7/SSH-2013-1 ATHEME 613465 (European Commission), ISQI 2011 JSH2 004 1 (Agence Nationale de la Recherche), FFI2011-29218, FF2011-26906, FFI2014-52015-P, FFI2017-82547-P (Ministerio de Ciencia e Innovación, Spanish Government).

## ACKNOWLEDGMENTS

We would like to thank Tasos Chatzikonstantinou and Ming Xiang for their help and comments. We are also grateful to the two reviewers of the paper, John Drury and Leticia Pablos for their insightful and careful reading of the paper. The many suggestions they made helped us sharpen our thoughts, and led to considerable improvements. Usual disclaimers apply.

- Drenhaus, H., Saddy, D., and Frisch, S. (2005). “Processing negative polarity items: When negation comes through the backdoor,” in *Linguistic Evidence—Empirical, Theoretical, and Computational Perspectives*, eds S. Kepser and M. Reis (Berlin; New York, NY: Mouton de Gruyter), 145–165.
- Duffley, P., and Larrivée, P. (2012). Collocation, interpretation and explanation: the case of JUST ANY. *Lingua* 122, 24–40. doi: 10.1016/j.lingua.2011.10.008
- Duffley, P., and Larrivée, P. (2015). A fresh look at the incompatibility between *any* and veridical contexts. The quality of indefiniteness is not strained. *Lingua* 158, 35–53. doi: 10.1016/j.lingua.2015.01.004
- Elordieta, G. (1997). “Accent, tone and intonation in lekeitio basque,” in *Issues in the Phonology and Morphology of the Major Iberian Languages*, eds M.-G. Fernando and M.-F. Alfonso (Washington, DC: Georgetown University Press), 4–78.
- Elordieta, G. (2003). “Intonation,” in *A Grammar of Basque*, eds J. I. Hualde and J. Ortiz de Urbina (Berlin: Mouton de Gruyter), 72–112.
- Elordieta, G., and Irurtzun, A. (2009). The relationship between meaning and intonation in nonexhaustive answers: evidence from Basque. *Linguist. Rev.* 3, 261–291.
- Ernst, T. (2009). Speaker oriented adverbs. *Nat. Lang. Linguist. Theory* 27, 497–544. doi: 10.1007/s11049-009-9069-1
- Etxeberria, U., and Irurtzun, A. (2004). “Prosodic features with semantic interpretation,” in *Proceedings of JEL 2004 Domains*, eds H. Demirdache, S. Waquier-Graveline, and O. Crouzet (Nantes: AAI), 95–101.
- Etxeberria, U., and Irurtzun, A. (2015). The emergence of scalar meanings. *Front. Psychol.* 6:141. doi: 10.3389/fpsyg.2015.00141
- Farkas, D. (1998). “Dependent indefinites,” in *Empirical Issues in Formal Syntax and Semantics*, eds F. Corblin, D. Godard, and J.-M. Marandin (Bern: Peter Lang Publishers), 243–268.
- Fauconnier, G. (1975). *Polarity and the Scale Principle*, Vol. 11. Chicago, IL: Chicago Linguistics Society.
- Friederici, A. D. (1995). The time course of syntactic activation during language processing: a model based on neuropsychological and neurophysiological data. *Brain Lang.* 50, 259–281. doi: 10.1006/brln.1995.1048
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84. doi: 10.1016/S1364-6613(00)01839-8

- Friederici, A. D., and Kotz, S. A. (2003). The brain basis of syntactic processes: functional imaging and lesion studies. *Neuroimage* 20, S8–S17. doi: 10.1016/j.neuroimage.2003.09.003
- Friederici, A. D., and Weissenborn, J. (2007). Mapping sentence form onto meaning: the syntax-semantic interface. *Brain Res.* 1146, 50–58. doi: 10.1016/j.brainres.2006.08.038
- Gajewski, J. (2002). *L-analyticity and Natural Language*, Cambridge, MA: MIT.
- Giannakidou, A. (1995). “Linking mood to polarity sensitivity: the case of the Modern Greek subjunctive,” in *Proceedings of CONSOLE 3*, eds A. Bisetti, J. Costa, R. Goedemans, N. Munaro, and R. van de Vijver (Venice: University of Venice), 71–83.
- Giannakidou, A. (1997). *The Landscape of Polarity Items*. Ph.D, University of Groningen.
- Giannakidou, A. (1998). *Polarity Sensitivity as (non)Veridical Dependency*. Amsterdam: John Benjamins.
- Giannakidou, A. (2000). Negative ... concord? *Nat. Lang. Linguist. Theory* 18, 457–523. doi: 10.1023/A:1006477315705
- Giannakidou, A. (2001). The meaning of free choice. *Linguist. Philos.* 24, 659–735. doi: 10.1023/A:1012758115458
- Giannakidou, A. (2006). Only, emotive factives, and the dual nature of polarity dependency. *Language (Baltim)* 82, 575–603. doi: 10.1353/lan.2006.0136
- Giannakidou, A. (2009). The dependency of the subjunctive revisited: temporal semantics and polarity. *Lingua* 119, 1883–1908. doi: 10.1016/j.lingua.2008.11.007
- Giannakidou, A. (2011). “Negative and positive polarity items,” in *Semantics: An International Handbook of Natural Language Meaning*, eds K. von Stechow, C. Maienborn, and P. Portner (Berlin: de Gruyter), 1660–1712.
- Giannakidou, A. (2016). “Evaluative subjunctive and non-veridicality,” in *Revisiting Mood, Aspect and Modality: What is a Linguistic Category*, eds J. Blaszczak, D. Anastasia Giannakidou, and K. Klimek-Jankowska (University of Chicago Press), 141–173. doi: 10.7208/chicago/9780226363660.003.0005
- Giannakidou, A. (2017). *Polarity in the Semantics of Natural Language*. Oxford Encyclopedia in Linguistics. Oxford: Oxford University Press.
- Giannakidou, A., and Cheng, L. (2006). (In)definiteness, polarity, and the role of wh-morphology in free choice. *J. Semant.* 23, 135–183. doi: 10.1093/jos/fll001
- Giannakidou, A., and Mari, A. (2018). A unified analysis of the future as epistemic modality: the view from Greek and Italian. *Nat. Lan. Linguist. Theory.* 36, 85–129. doi: 10.1007/s11049-017-9366-z
- Giannakidou, A., and Quer, J. (2013). Exhaustive and non-exhaustive variation with anti-specific indefinites: free choice and referential vagueness in Greek, Catalan, and Spanish. *Lingua* 26, 120–149. doi: 10.1016/j.lingua.2012.12.005
- Giannakidou, A., and Yoon, S. (2016). Scalar marking without scalar meaning: non-scalar, non-exhaustive NPIs in Greek and Korean. *Language (Baltim)* 92, 522–556. doi: 10.1353/lan.2016.0047
- Giannakidou, A., and Zeijlstra, H. (2017). “The landscape of negative dependencies: negative concord and n-words,” in *The Blackwell Companion to Syntax*, 2nd Edn., Eds M. Everaert and H. van Riemsdijk. doi: 10.1002/9781118358733.wbsyncom10
- Gouvêa, A. C., Phillips, C., Kazanina, N., and Poeppel, D. (2010). The linguistic processes underlying the P600. *Lang. Cogn. Process.* 25, 149–188. doi: 10.1080/01690960902965951
- Grano, T. (2011). “Mental action and event structure in the semantics of try,” in *Semantics and Linguistic Theory*, eds A. Chereches and D. Lutz (New Brunswick, NJ: Rutgers University), 426–443.
- Hagoort, P., Brown, C. M., and Groothusen, J. (1993). The syntactic positive shift as an ERP measure of syntactic processing. *Lang. Cogn. Process.* 8, 439–483.
- Hoeksema, J. (2010). Dutch ENIG: from nonveridicality to downward entailment. *Nat. Lang. Linguist. Theory* 28, 837–859. doi: 10.1007/s11049-010-9110-4
- Horn, L. (1972). *On the Semantic Properties of Logical Operators in English*, Ph.D. UCLA.
- Horn, L. (1989). *A Natural History of Negation*. Stanford, CA: CSLI Publications.
- Horn, L. (2002). “Assertoric inertia and NPI licensing,” in *Proceedings of the Annual Meeting of the Chicago Linguistic Society* (Chicago, IL: University of Chicago), 38.
- Hualde, J. I., and Elordieta, G. (2014). “Intonation in Basque,” in *Prosodic Typology II: The Phonology of Intonation and Phrasing*, ed S.-A. Jun (Oxford: Oxford University Press), 405–463.
- Iruztun, A. (2013). Focal NSR and stress placement in Basque phrases and N+N compounds. *Linguist. Anal.* 38, 207–242.
- Israel, M. (1996). Polarity sensitivity as lexical semantics. *Linguist. Philos.* 19, 619–666.
- Israel, M. (2011). *The Grammar of Polarity: Pragmatics, Sensitivity, and the Logic of Scales*. Cambridge: Cambridge University Press.
- Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.
- Jackson, S. (2007). *Information, Truth, Structure, and Sound*. Ph.D, University of Arizona.
- Kaan, E., Harris, A., Gibson, E., and Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Lang. Cogn. Process.* 15, 159–201. doi: 10.1080/016909600386084
- Kadmon, N., and Landman, F. (1993). ‘Any’. *Linguist. Philos.* 16, 353–422. doi: 10.1007/BF00985272
- Kennelly, S. D. (2003). The implication of quantification for the role of focus in discourse structure. *Lingua* 113, 1005–1088. doi: 10.1016/S0024-3841(03)00013-5
- Kennelly, S. D. (2004). *Quantificational Dependencies*. Ph.D. Dissertation, University of Utrecht.
- Klima, E. S. (1964). “Negation in English,” in *The Structure of Language*, eds J. A. Fodor and J. J. Katz (Englewood Cliffs, NJ: Prentice-Hall), 246–323.
- Krifka, M. (1995). The semantics and pragmatics of polarity items in assertion. *Linguist. Anal.* 15, 209–257.
- Krifka, M. (1998). Scope inversion under the rise-fall contour in German. *Linguist. Inq.* 29, 75–112. doi: 10.1162/002438998553662
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Res.* 1146, 23–49. doi: 10.1016/j.brainres.2006.12.063
- Kuperberg, G. R. (2013). “The pro-active comprehender: what event-related potentials tell us about the dynamics of reading comprehension,” in *Unraveling the Behavioral, Neurobiological, and Genetic Components of Reading Comprehension*, eds B. Miller, L. Cutting, and P. McCardle (Baltimore: Paul Brookes Publishing), 176–192.
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Ladusaw, W. A. (1980). *Negative Polarity Items as Inherent Scope Relations*. Ph.D, Dissertation, University of Texas at Austin.
- Laka, I. (1994). *On the Syntax of Negation. Outstanding Dissertations in Linguistics Series*, New York, NY; London: Garland Publishing Co.
- Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: [de]constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532
- Li, F., González-Fuente, S., Pietro, P., and Espinal, M. T. (2016). Is mandarin chinese a truth-based language? Rejecting responses to negative assertions and questions. *Front. Psychol.* 7:1967. doi: 10.3389/fpsyg.2016.01967
- Lin, J. (2015). *On the Acquisition of Negative Polarity Items*. Ph.D. University of Amsterdam.
- Linebarger, M. (1987). Negative polarity and grammatical representation. *Linguist. Philos.* 10, 325–387. doi: 10.1007/BF00584131
- Marti, L. (2001). “Intonation, scope and restriction on quantifiers,” in *Proceedings of WCCFL 20*, eds K. Megerdooian and L. Bar-el (Somerville, MA: Cascadia Press), 372–385.
- Menéndez-Benito, P. (2010). On universal free choice items. *Nat. Lang. Semant.* 18, 33–64. doi: 10.1007/s11050-009-9050-x
- Milsark, G. (1977). Toward an explanation of certain peculiarities of the existential construction in English. *Linguist. Anal.* 3, 1–28.
- Osterhout, L., and Holcomb, P. J. (1992). Event-related potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806. doi: 10.1016/0749-596X(92)90039-Z
- Osterhout, L., Holcomb, P. J., and Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 786–803. doi: 10.1037/0278-7393.20.4.786
- Partee, B. H. (1988). “Many quantifiers,” in *Proceedings of Fifth ESCOL*, eds J. Powers and K. de Jong (Columbus, OH: The Ohio State University), 383–402.

- Phillips, C., Kazanina, N., and Abada, S. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cogn. Brain Res.* 22, 407–428. doi: 10.1016/j.cogbrainres.2004.09.012
- Progovac, L. (1994). *Positive and Negative Polarity: A Binding Approach*. Cambridge: Cambridge University Press.
- Progovac, L. (2005). “Negative and positive feature checking and the distribution of polarity items,” in *Negation in Slavic*, eds S. Brown and A. Przepiorkowski (Bloomington, IN: Slavica Publishers), 179–217.
- Quer, J. (1998). *Mood at the Interface*. Ph.D. thesis. University of Utrecht.
- Saddy, D., Drenhaus, H., and Frisch, S. (2004). Processing polarity items: contrasting licensing costs. *Brain Lang.* 90, 495–502. doi: 10.1016/S0093-934X(03)00470-X
- Shao, J., and Neville, H. (1998). Analyzing semantic processing using event-related potentials. *Newsl. Center Res. Lang.* 11, 3–20.
- Simons, M., David, B., Craige, R., and Judith, T. (2017). The best question: explaining the projection behavior of factives. *Discourse Process.* 54, 187–206. doi: 10.1080/0163853X.2016.1150660
- Steinhauer, K., and Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain Lang.* 120, 135–162. doi: 10.1016/j.bandl.2011.07.001
- Steinhauer, S., Drury, J., Portner, P., Walenski, M., and Ullman, M. (2010). Syntax, concepts, and logic in the temporal dynamics of language comprehension: evidence from event-related potentials. *Neuropsychologia*, 48, 1525–1542. doi: 10.1016/j.neuropsychologia.2010.01.013
- Tesan, G., Johnson, B. W., and Crain, S. (2012). How the brain responds to any: an MEG study. *Brain Lang.* 120, 66–72. doi: 10.1016/j.bandl.2011.08.006
- Tonhäuser, J. (2016). “Prosodic cues to presupposition projection,” in *Semantics and Linguistic Theory (SALT)*, Vol. 26, eds C. J. M. Carol-Rose, M. Little, and D. Burgdorf (Austin, TX: University of Texas), 934–960.
- Tsimpli, I. M., and Roussou, A. (1996). Negation and polarity items in Modern Greek. *Linguist. Rev.* 13, 1–33. doi: 10.1515/tlir.1996.13.1.49
- Van Petten, C., and Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *Int. J. Psychophysiol.* 83, 176–190. doi: 10.1016/j.ijpsycho.2011.09.015
- Veloudis, I. (1982). *Negation in Modern Greek*. Ph.D. University of Reading.
- von Stechow, K. (1999). NPI-licensing, strawson-entailment, and context-dependency. *J. Semant.* 16, 97–148. doi: 10.1093/jos/16.2.97
- Xiang, M., Dillon, B., and Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain Lang.* 108, 40–55. doi: 10.1016/j.bandl.2008.10.002
- Xiang, M., Grove, J. and Giannakidou, A. (2013). Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Front. Psychol.* 4:708. doi: 10.3389/fpsyg.2013.00708
- Xiang, M., Grove, J., and Giannakidou, A. (2016). Semantic and pragmatic processes in the comprehension of negation: an event related potential study of negative polarity items. *J. Neurolinguistics* 38, 71–88. doi: 10.1016/j.jneuroling.2015.11.001
- Yanilmaz, A., and Drury, J. E. (2018). Prospective NPI licensing and intrusion in Turkish. *Lang. Cogn. Neurosci.* 33, 111–138. doi: 10.1080/23273798.2017.1371779
- Yurchenko, A., den Ouden, D. B., Hoeksema, J., Dragoy, O., Hoeks, J., Stowe, L., et al. (2013). Processing polarity: ERP evidence for differences between positive and negative polarity. *Neuropsychologia* 5, 132–141. doi: 10.1016/j.neuropsychologia.2012.10.028
- Zeijslstra, H. (2004). *Sentential Negation and Negative Concord*. Ph.D. University of Amsterdam.
- Zwarts, F. (1995). Nonveridical con-texts. *Linguist. Anal.* 25, 286–312.
- Zwarts, F. (1996). “A hierarchy of negative expressions” in *Negation: A Notion in Focus*, ed H. Wansing (Berlin: de Gruyter), 169–194.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Giannakidou and Etxeberria. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Handling Sign Language Data: The Impact of Modality

Josep Quer<sup>1\*</sup> and Markus Steinbach<sup>2</sup>

<sup>1</sup>ICREA-Pompeu Fabra University, Barcelona, Spain, <sup>2</sup>University of Göttingen, Göttingen, Germany

Natural languages come in two different modalities. The impact of modality on the grammatical structure and linguistic theory has been discussed at great length in the last 20 years. By contrast, the impact of modality on linguistic data elicitation and collection, corpus studies, and experimental (psycholinguistic) studies is still underinvestigated. In this article, we address specific challenges that arise in judgment data elicitation and experimental studies of sign languages. These challenges are related to the socio-linguistic status of the Deaf community and the larger variability across signers within the same community, to the social status of sign languages, to properties of the visual-gestural modality and its interface with gesture, to methodological aspects of handling sign language data, and to specific linguistic features of sign languages. While some of these challenges also pertain to (some varieties of) spoken languages, other challenges are more modality-specific. The special combination of the challenges discussed in this article seems to be a specific facet empirical research on sign languages is faced with. In addition, we discuss the complementarity of theoretical approaches and experimental studies and show how the interaction of both approaches contributes to a better understanding of sign languages in particular and linguistic structures in general.

**Keywords:** sign language, native signers, language modality, language analysis, language documentation, data collection

## OPEN ACCESS

### Edited by:

Ángel J. Gallego,  
Autonomous University of Barcelona,  
Spain

### Reviewed by:

Evelina Leivada,  
UiT The Arctic University of Norway,  
Norway

Francesca Peressotti,  
University of Padova, Italy

### \*Correspondence:

Josep Quer  
josep.quer@upf.edu

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 07 May 2018

**Accepted:** 19 February 2019

**Published:** 12 March 2019

### Citation:

Quer J and Steinbach M (2019)  
Handling Sign Language  
Data: The Impact of Modality.  
*Front. Psychol.* 10:483.  
doi: 10.3389/fpsyg.2019.00483

## INTRODUCTION

Sign and spoken languages use two different modalities, the visual-gestural modality of sign languages and the oral-auditory modality of spoken languages. Although the two modalities clearly differ in the production and perception of communicative signals, the underlying linguistic structures seem to be very similar across both modalities (Meier, 2002, 2012; Sandler and Lillo-Martin, 2006).<sup>1</sup> In addition, psycho- and neurolinguistic studies with non-impaired and impaired deaf signers show that sign languages access the same neural networks involved in auditory speech processing, albeit with some concrete modality-specific features (Poizner et al., 1987; Emmorey, 2002, 2003; Corina and Knapp, 2006; Campbell et al., 2008; Corina and Spotswood, 2012; Dye, 2012; Woll, 2012). Nevertheless, sign languages retain some modality-specific properties that may impact the linguistic structure and the cognitive processes underlying the perception and production of signed communication and that have an influence on the handling of sign language data (cf. van Herreweghe and Vermeerbergen, 2012; Orfanidou et al., 2015). First of

<sup>1</sup>This view is not fully shared by some sign language scholars working in cognitive linguistics, who put the emphasis on the differences derived from the visual-gestural modality (cf. Liddell, 2003, a.o.).

all, sign languages employ various articulators such as the hands, the upper part of the body, the head, and the face to express grammatical features simultaneously. Second, sign languages use the geometrical properties of the signing space to realize morphosyntactic, semantic, and pragmatic categories in the three-dimensional signing space (Engberg-Pedersen, 1993; Padden, 1998; Aronoff et al., 2005; Pfau and Steinbach, 2016; Steinbach and Onea, 2016). Third, sign languages grammaticalize and integrate gestural elements, since sign languages and manual as well as non-manual gesture use the same modality. As a consequence, the interface between these two systems is permeable (Liddell and Metzger, 1998; Emmorey, 1999; Liddell, 2003; Pfau and Steinbach, 2011; Grosvald et al., 2012; Goldin-Meadow and Brentari, 2017) and leads to a more prominent presence of iconicity at different grammatical levels (Taub, 2012). By contrast, there is much less transparency between the signals used in auditory communication and their meaning (Schlenker, 2018).

Besides these linguistic differences, sign languages differ from many spoken languages also in various socio-linguistic dimensions (Aronoff et al., 2005). In the next section, we first deal with these dimensions (the data source problem) and discuss consequences of the quite heterogeneous group of sign language users for linguistic studies. In the second part, we turn to the impact of modality on the elicitation and annotation of sign language data. Specific practical and conceptual challenges may arise from the heterogeneity of linguistic informants and subjects, the lack of a writing system, the material properties of the data, and the modality-specific linguistic aspects of sign languages mentioned above. Note that while some of these challenges may also hold true for some varieties of spoken languages (e.g., for spoken languages without a written form used by linguistic minorities in multilingual contexts), other challenges are clearly modality-specific. Since the focus of this article is on sign language data handling, we discuss spoken languages only in passing. It will, however, turn out that the expertise gained in linguistic research on sign languages paves the way for new multimodal investigations of spoken languages.

## THE DATA SOURCE PROBLEM

Formal linguistic analysis typically relies on evidence provided by native speakers of the language or variety under study. This can involve different types of collected spontaneous or semi-spontaneous productions, elicited utterances, or grammaticality judgments. Despite the unavoidable abstraction across different speakers, it is taken for granted that their competence is similar enough by virtue of having acquired the language natively in a typical, unproblematic fashion.<sup>2</sup> However, such a simple assumption cannot be made for sign languages because of their highly idiosyncratic sociolinguistic settings and in particular their dominant acquisition patterns (Schembri and Lucas, 2015).

<sup>2</sup>This amounts to Labov's Consensus Principle: "If there is no reason to think otherwise, assume that the judgments of any native speaker are characteristic of all speakers of the language" (Labov, 1996).

At least for Western societies, it is often taken for granted that only 5–10% of deaf children are born to deaf parents or in an environment where there is adequate sign language input for the child to develop language competence in a natural way (Neidle et al., 2000; Mitchell and Karchmer, 2004). This means that most deaf babies (the remaining 90–95% of deaf children at birth) are not surrounded by a natural language in the visual-gestural modality, which is fully accessible to them, but rather by spoken language. A variety of factors determines the language acquisition path for them: (1) hearing parents can decide to learn and use sign language themselves with the child (a very small percentage, cf. Chen-Pichler and Lillo-Martin, 2018); (2) parents can choose a schooling model that favors interaction and instruction in sign language to different degrees (from deaf schools to bilingual programs embedded in regular schools); (3) parents are often confronted with the choice of giving their child a cochlear implant that will facilitate access to the spoken language signal after regular and intensive training. These elements already make it evident that for most deaf children access to language during the critical period will be uncertain, to say the least, and in any event more incomplete or degraded than in the default case where rich language input is part of the environment. Take for instance the favorable, albeit uncommon, case where parents decide to use sign language with the child and choose for a day care and school that offers a bimodal bilingual approach: even in this favorable case, most adult language models will be non-native (hearing parents, hearing teachers and classroom interpreters that learn sign language as a second language) and some of them will use mixed forms of language (in general, spoken structure imposed on sign), thus providing an input that is strictly speaking qualitatively different from the native one. The obvious consequence of this situation is that the majority of signers in Deaf communities have acquired their sign language under such special circumstances and do not fall under the strict definition of native speakers or signers. To this we must add the fact that regular contact with sign language may happen at different stages in life and it is quite common for deaf children to be initially raised only with spoken language and for them to be exposed to sign language past the first year of life, turning them technically into early or late learners of a sign language is not L2 learning, but simply delayed L1 learning at an abnormal age (late childhood, adolescence, or adulthood), leading to abnormal neurological mappings of language (Mayberry, 2010; Mayberry and Kluender, 2017; Woll, 2018). Research has confirmed the expectation that such different paths of language acquisition should impact on language competence (Boudreault and Mayberry, 2006; Cormier et al., 2012; Skotara et al., 2012; Hänel-Faulhaber et al., 2014, 2018, unpublished; Lillo-Martin, 2018).

Next to such atypical language acquisition paths, linguistic research must also take into account that most deaf signers have bilingual competence as a result of spoken language acquisition to varying degrees, even if it is the language acquired first chronologically. Nowadays spoken language competence

in signers takes two different paths: mostly competence in the written form, as a result of schooling and interaction with the ambient hearing society; competence in the spoken modality because of the spreading of cochlear implants, which typically involves mainstreaming in education and intensive speech therapy. In this picture, postlingual deaf children constitute yet another case, as they will have acquired spoken language for the most part when they lose their hearing, thus being able to rely on full-fledged language acquisition during the first year of life as base for subsequent sign language acquisition.

Among bilingual signers, another group must be taken into account: hearing native signers, most commonly known as bimodal bilinguals (Branchini and Donati, 2016; Emmorey et al., 2016; Lillo-Martin et al., 2016). This population is formed by hearing children of Deaf adults (CODAs) who have been exposed to sign from birth and have acquired it natively while acquiring the ambient spoken language at the same time in the larger family context, at school and in social interaction. CODAs form an idiosyncratic language group that has only received attention quite recently within the study of bilingual competence. In a sense, they represent the unique case of full simultaneous bilingualism in two modalities, given their unproblematic access to the input in both sign and spoken language. They offer a unique window into the bilingual mind that can process and externalize utterances realized in the two channels simultaneously, namely code blends. As for their sign language competence, it has been paralleled to that of heritage speakers, since they will use it only in family or community contexts and will use the ambient spoken language most of the time (unless they become sign language interpreters, of course, or have deaf children or a deaf spouse) (Quadros, 2018).

This cursory description of the factors that impact on the individual competence of sign language users highlights the complexity of trying to characterize language competence across a signing community. It is still common practice – among formal linguists at least – to study sign language structure relying on evidence provided by native signers, even though they constitute a very small minority within signing communities. Their scarcity often involves difficulties in accessing native signers as language consultants that are willing to collaborate and provide data, and in some cases, it cannot even be feasible, as discussed by Costello et al. (2008). The situation might be even more problematic if the usually quoted rates of deaf-of-deaf individuals are in reality lower in countries other than the United States, as argued by Johnston (2006).

Given these limitations, some alternatives have been proposed. One of them consists in working with consultants that get as close as possible to a native signer, as put forth in Mathur and Rathmann (2006): (1) exposure to a sign language by the age of 3; (2) daily contact with a sign language in the Deaf community for longer than 10 years. For linguistic research, they also required (3) capability to make grammaticality judgments with ease. Freel et al. (2011) also establish this age limit of 3 in the acquisition of sign language in order to count someone as native signer. Such accommodations seem desirable in practical terms, but it might be the case that even with these slight departures

from strict nativehood, it is still hard to find sign language consultants, given their scarcity in some areas.

An obvious reaction to the difficulty of working with native signers to obtain fresh data would be to resort to existing resources such as grammars and corpora. Unfortunately, such tools do not exist for most sign languages. With a few exceptions, reference works or even partial descriptions of grammar components (phonology, morphology, syntax, etc.) are lacking. An attempt to remedy this situation has been undertaken by developing a detailed guide to sign language grammar writing, the *SignGram Blueprint* (Quer et al., 2017), which will also be implemented as an online grammar writing tool on the platform currently developed by the SIGN-HUB project (H2020: 2016–2020). By the end of this project, the grammars of six languages will be available, and hopefully, this step will set the trend for other sign languages and steadily fill the vast gap that we are currently faced with in terms of background grammatical information for languages in the visual-gestural modality.

Sign language corpora are not available as a default (e.g., there is no reference corpus for American Sign Language (ASL) despite being the longest studied sign language), but different projects in Europe and Australia have addressed this need and developed representative corpora for certain sign languages that gather spontaneous or semi-spontaneous signing on the basis of different tasks or elicitation techniques.<sup>3</sup> Some of them are already available, while others are currently being built. But even if a corpus is available, one general problem of most corpora is that they lack detailed linguistic annotation, especially at the levels of morphosyntax and (discourse) semantics. Hence, they can be used for linguistic investigations only to a limited extent. A more general problem is the significance of corpora. Although corpus data are useful for the description of grammatical structures and sociolinguistic variation, they are known to be problematic for theoretical analysis, given the limitation that no negative evidence can be obtained (non-appearance in the corpus cannot be equated to ungrammaticality). In the case of sign languages, the individual variation referred to above must be added to the complications of relying on corpus data. The issue can be mitigated, thanks to the use of metadata about the consultants recorded in such a way that one could in principle select only production by signers with a common linguistic profile (e.g., natives). However, the best situation will be one in which data types can be combined, for instance, by collecting corpus data and eliciting grammaticality judgments. Another technique used in sign language research is to discuss data with consultants, whether they have been produced by themselves (and played after an acceptable time lapse) or by others (as with corpus data, for instance).

Having access to native signers as consultants or enough relevant corpus or elicited data, though, is not enough to be able to guarantee that we are researching a particular sign language. As is well known from spoken language research, variation within a linguistic domain needs to be taken into account when defining the object of study. A similar

<sup>3</sup>For an overview, see <https://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/sl-corpora.html>.

situation arises with sign language data but sometimes with parameters of variation that are unique to the visual-gestural modality.

Geographical variation is certainly also present in sign language communities, but with some idiosyncratic features vis-à-vis spoken languages. Till quite recently, regional variants of sign languages were only indirectly determined by geographical area of use: given the dispersal of deaf individuals within hearing societies, the two poles of emergence and irradiation of signed varieties were mainly: (1) deaf (boarding) schools and (2) deaf clubs. These institutions created contexts where deaf signers formed a critical mass for language use, but crucially also for language acquisition/learning. The impact of educational institutions on variation is clear in many countries, as in the Netherlands, where five regional variants can be identified as a consequence of the existence of five different deaf schools (Schermer, 2012). This kind of variation mainly affects the lexicon (especially in certain lexical domains like numerals, names of weekdays and months, colors, and kinship terms; gender differences can even be traced back to the existence of segregated schooling), phonology, and grammar.

There are very few studies that focus on variation from a formal point of view, but the potential of corpus data analysis from this perspective is clear. One of them deals with the position of *wh*-elements across regional variants of Italian Sign Language (LIS) (Geraci et al., 2015), and it interestingly concludes that a variable like age (and, linked to that, language awareness) plays a decisive role in the position of *wh*-elements. With this, we see how sociolinguistic factors such as schooling, language contact and awareness can determine language production (and arguably competence). Another interesting example of research that targets grammatical phenomena relying on data that reflect variation concerns the syntactic position of the agreement auxiliary PAM in German Sign Language (DGS) (Macht and Steinbach, 2018): a rough partition of the DGS domain in north, west, east, and south shows that in the former three the preverbal realization of the auxiliary is clearly predominant (72% up to 85% of the instances), while in the south, it appears before and after the predicate in almost the same percentages. These results clearly point to a different syntactic derivation across areas of the same language domain that need to be further investigated with respect to other structural phenomena.

With this brief overview of the individual and social factors that can determine language competence in signers it becomes evident that data elicitation, grammaticality judgments tasks and experimental studies should be carried out with particular care in order to reach reliable generalizations about a particular sign language.

## MODALITY AND DATA COLLECTION

In the previous section, we discussed various individual and social factors that may affect any kind of empirical and

experimental data collection, annotation, evaluation, and documentation. Some of them are related to the fact that sign languages are minority languages and that deaf native signers form a unique linguistic minority. Others are related to the specific kind of language acquisition, the influence of the ambient spoken language(s), and the (linguistic) heterogeneity of the Deaf community. Before we turn to modality-specific aspects that may have an impact on data collection, we briefly discuss how these aspects need to be considered in empirical investigations of sign languages (for more details, see Orfanidou et al., 2015).

First of all, working with linguistic minorities requires the strict compliance of the highest standards of ethical principles. This is especially important since most kinds of data collection involve video recording, which means that informants or subjects are always visible and clearly identifiable.<sup>4</sup> Because of the very specific properties of visual-gestural languages, data cannot easily be made anonymous since each part of the upper part of the body and the face conveys important grammatical and pragmatic information. This brings us to the second aspect: Sign language data are typically video data, that is, sign language linguists always use, collect, annotate, and analyze quite complex visual information. As opposed to many spoken languages, sign languages do not have a written form that can be used for data collection and data storage. Linguistic glosses used in research on sign languages are always only simplified linguistic representations of the multidimensional visual information of a corresponding video documenting the utterance (Frishberg et al., 2012; Crasborn, 2015). We will come back to this issue below. Moreover, effective tools for automatic processing and annotation of sign language video data are not available yet (see Hanke, 2016 and below). Third, a careful collection of metadata is inevitable to specify the significance of a specific set of data collected in an empirical study. The validity of data depends on the kind of informants and subjects involved in the study. A related fourth aspect is that each empirical study should start with a clear definition of the socio-linguistic features of informants and subjects of a study to get optimal and valuable empirical data for the linguistic research question under discussion. This is especially important for studies with smaller groups of informants and subjects. Fifth, empirical studies should always be conducted in a sign language-friendly environment, which includes interaction and instruction in sign language, and use deaf friendly research methods. Ideally, the study is conducted by a deaf researcher. Likewise, the data should be annotated and evaluated by mixed teams including deaf researchers. And last but not least, linguists should be aware of the fact that sign language users are not only a linguistic minority but are in many countries and regions also very small groups with many

<sup>4</sup>Two groups deserve closer attention, namely Deaf children and individuals with impairments in sign languages, which usually neither receive assessment nor intervention (for more information on ethics issues, see also Baker (2012) and the ethics statement of the Sign Language Linguistics Society: <http://slls.eu/slls-ethics-statement/>)

non-academic members. Therefore, any kind of data collection should respect the specific needs of these groups and include regular activities of transfer of knowledge and dissemination in the local sign language.

The heterogeneity of the Deaf community may also directly affect the results of linguistic studies. In many Western societies, sign languages have been recognized only recently. Therefore, informants and subjects may have grown up in bilingual or strict oralist environments where sign languages have not been taught at school (and have even not been used in the classroom). This situation—which is not modality-specific but typical for many sign languages—can have an influence on the evaluation of linguistic data and grammaticality judgments of signers, especially in tasks where certain information (such as, for instance, linguistic contexts) is provided in a spoken language or where answers to linguistic questions have to be given in a spoken language. Like for other bilinguals, it has been shown that deaf bimodal bilinguals also activate the second language (i.e., the ambient spoken language) while processing the first language (i.e., the native sign language) (Hosemann, 2015). Therefore, the specific language awareness in oral environments and the fact that deaf signers are typically bilingual should be taken into account, as mentioned earlier. In general, using spoken language input for elicitation tasks should be avoided if possible in order to minimize interference (Nishio et al., 2010). This means that instructions and contexts necessary for controlled data elicitation have to be provided in the sign language under investigation (for the importance of controlled data elicitation, see Matthewson, 2004).

Let us now turn to modality-specific aspects of sign languages that are relevant in this context. First of all, unlike spoken languages, sign languages are characterized by a relatively long transition phase between two linguistic units (compared to spoken languages).<sup>5</sup> One reason for this is that sign languages, unlike spoken languages, make use of relatively massive articulators that execute long movements (Meier, 2002, 2012). Consequently, phonological parameters change much slower than in spoken languages. In addition, phonological parameters can be realized simultaneously, that is, in the transition phase more than one parameter may change at the same time. Hence, the transition phase is not linguistically empty but already contains a lot of linguistic information (change of handshape, direction of movement, etc.) that can be used to identify the upcoming sign (Emmorey and Corina, 1990; Hosemann et al., 2013), and thus raises some conceptual and practical issues for empirical studies and corpus linguistics. Let us briefly discuss three problems here: (1) In sign languages, the

presentation of complex stimuli sign by sign is problematic since the hands must either go back to a neutral position in the signing space or the videos are cut in the middle of the transition phase. Both options are highly artificial because the transition phase connecting two signs is missing or interrupted. Moreover, additional non-manual markers may simultaneously scope over more than one sign, which makes a presentation of complex stimuli sign by sign even more unnatural. (2) In corpus annotation, we are faced with the problem of identifying the sign on- and offset (Hanke et al., 2012). A too strict definition of on- and offset would leave us with a lot of intermediate material, the transition phase, that does not have any linguistic value. A flexible definition leaves us with the problem that on- and offset can only be identified in context and may vary between examples and annotators. In both cases, this may falsify the results of statistical evaluations of corpus data. (3) In online studies, the identification of the sign onset directly affects the time-locked evaluation of the experimental data. However, for the data evaluation, the experimenter must decide which point in time s/he uses to identify the onset of a sign. Things are even more complex since the recognition of the onset of a sign by the subjects may vary from experiment to experiment. The recognition of an upcoming event (i.e., a sign) can depend on information available in context, on information provided simultaneously by manual and non-manual activities and on properties of the critical sign itself. Therefore, the experimenter should handle this problem carefully and transparently. A related practical aspect is that in corpora and experiments, the sign onset and the trigger (i.e., the time-locked position) have to be identified manually in the videos. This means that sign language competent annotators have to determine the relevant points in time in each video frame by frame, which is a highly time-consuming task (for a discussion of trigger identification in ERP experiments, cf. Hosemann et al., 2013, 2018).

Video stimuli pose yet another challenge for another kind of experimental online studies, namely eye tracking studies. In eye-tracking experiments on sign languages, typical measures such as fixation and saccades are more difficult to define and to relate to the linguistic input than in typical eye-tracking experiments on spoken languages that present the input in written form. This might be one reason why up to now only a few eye tracking studies on sign languages have been conducted. These studies focus either on eye gaze of signers during production (Thompson et al., 2006, 2009; Hosemann, 2011) or on the question whether the addressee typically focuses on the face of the signer (Muir and Richardson, 2005; Emmorey et al., 2009). Very few studies conducted a visual world experiment where the visually presented items are not linguistic objects (e.g., individual signs or complex utterances) but pictures somehow related to the linguistic input (Thompson et al., 2013; Lieberman et al., 2015, 2017). Hence, on the one hand, the lack of a writing system prevents the linguistic study of eye movements during processing the written form of a language. This means that standard methods, which are well established in psycholinguistic research on the written form of spoken

<sup>5</sup>One reviewer mentioned that even in spoken languages, onset time is not always strictly accurate. While this is definitely true, the determination of the sign onset poses quantitatively and qualitatively different (and more serious) problems in sign languages as compared to spoken languages. In addition, the development of automatic sign processing and parsing is still in the very early stage, which means that research on sign languages, unlike research on spoken languages, will require a lot of manual annotation even in the near future (cf. Sáfár and Glauert, 2012; Hanke, 2016 and the discussion below).

languages, cannot be applied to sign languages.<sup>6</sup> On the other, the presentation of visual stimuli (i.e., videos of naturally signed stimuli) makes the definition of areas of interests over many different stimuli and the linguistic evaluation of additional eye movements related to the linguistic input (e.g., in a visual world paradigm) more difficult. Hence, specific properties of the visual-gestural modality complicate the applicability of a standard online technique of experimental linguistic research.

Let us finally turn to the impact of the three modality-specific properties mentioned in the introduction: simultaneity, space, and gesture. All three properties require smart theoretical decisions and they cause extra effort in the transcription and annotation of linguistic examples collected in a corpus or in a production study (Frishberg et al., 2012; Orfanidou et al., 2015). On the one hand, the form and function of simultaneously used articulators need to be annotated on different tiers. Since these articulators express grammatical properties at different linguistic levels (prosody, morphology, syntax, semantics, and pragmatics) and interact in non-trivial ways, even simple examples require complex annotations (for a discussion of the annotation of action role shift, cf. e.g., Cormier et al., 2015). Given the fact that automatic segmentation and annotation are not available for sign language data yet, it is obvious that the complex annotation sequences of sign language data is extremely time consuming. A similar problem is the mapping of three-dimensional properties of the signing space onto a two-dimensional linguistic annotation schema. This does not only concern phonological properties of lexical signs but also grammatical features realized in the signing space. On the other hand, manual and non-manual gestures and signs are not always easy to distinguish (Goldin-Meadow and Brentari, 2017). This leads to the modality-specific problem to integrate gestures or gesture-like elements at various levels into the linguistic annotation. This problem presupposes, however, a clear theoretical definition of gesture and sign as well as the interaction of gestures and signs.

The three modality-specific properties also raise interesting questions for experimental studies and make cross-modal comparisons between spoken and sign languages difficult. Let us consider simultaneity and space first. Since sign languages use spatial and simultaneous markers to realize grammatical features, the creation of controlled stimuli is not always easy. Spatial grammatical features such as R(eferential)-loci can, for instance, be marked manually (movement and orientation of agreement verbs) or non-manually (body lean, head movement, or eye gaze). Experimental studies on the use of R-loci, may, for example, require the control of simultaneous non-manual identification of R-loci in the stimuli to investigate the manual grammatical properties of pronouns or agreement verbs. Hence, the experimenter may decide to neutralize the non-manual markers

in the examples. This may, however, result in quite unnatural stimuli and thus affect the results of the experiment (cf. Hosemann et al., 2018; Wienholz et al., 2018). The same holds true for other non-manuals such as mouthing or facial expressions. A related problem is that spatial features cannot be neutralized completely since any sign is produced in space. Therefore, even the production of a simple sign may affect spatial interpretations. By contrast, if we only use natural stimuli in experimental studies, we may not be able to control the stimuli to the extent necessary for a valid and reliable evaluation of the data. A similar problem can arise from the use of iconic signs and gestural elements in sign language, which may affect grammaticality judgments and trigger different paths of processing.

## SUMMARY

In this article, we have shown that sign language linguists are faced with a number of challenges that are either related to socio-linguistic aspects of the signing community (the data source problem) or to specific linguistic aspects of the visual-gestural modality and to methodological problems of sign language data collection, annotation, and stimuli creation (modality and data collection). In addition, we have argued that while some of these challenges also concern linguistic studies of spoken languages (particularly, of spoken varieties of small communities with no written tradition, such as in the so-called Italian dialects), other challenges are more modality-specific. Therefore, studies on sign languages are typically much more time-consuming than comparable investigations of spoken languages, especially of well-established and well-documented spoken languages. However, facing these challenges is worth the effort, since the expertise gained in empirical and experimental studies of sign languages and sign language documentation (reference grammars and corpora), while germane in several respects to empirical research in small spoken language communities is in other respects pioneering work and will pave the way for future multimodal investigations of spoken languages including co-speech gestures and facial expressions.

## AUTHOR CONTRIBUTIONS

Both authors JQ and MS have contributed equally to this article.

## FUNDING

This contribution has been made possible, thanks to the SIGN-HUB project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 693349. JQ contribution is also made possible by the Spanish Ministry of Economy, Industry, and Competitiveness and FEDER Funds (FFI2015-68594-P) and by the Government of the Generalitat de Catalunya (2017 SGR 1478). MS contribution is made possible by the German Science foundation (DFG-AZ: STE 958/8-1).

<sup>6</sup>This is, of course, also true for the processing of auditory stimuli. Note, however, that psycholinguistic studies on many (but not all) spoken languages can use two different input modalities (i.e., written and spoken modality) to investigate linguistic structure. By contrast, psycholinguistic studies on sign languages cannot draw on written stimuli. This makes a big difference for psycholinguistic investigations. The huge amount of psycholinguistic research on written language shows that written stimuli can successfully be used to get insight in the processing of spoken languages in general (although the written modality is not a simple copy of the spoken modality).

## REFERENCES

- Aronoff, M., Meir, I., and Sandler, W. (2005). The paradox of sign language morphology. *Language* 81, 301–344. doi: 10.1353/lan.2005.0043
- Boudreau, P., and Mayberry, R. I. (2006). Grammatical processing in American sign language: age of first-language acquisition effects in relation to syntactic structure. *Lang. Cogn. Process.* 21, 608–635. doi: 10.1080/01690960500139363
- Baker, A. (2012). “Ethics issues in sign language acquisition data archiving” in *IPROSLA: Integrating and publishing resources of language acquisition*. Available at [https://slls.eu/wp-content/uploads/2012/12/Ethics-issues-in-sign-language-acquisition-data-archiving-IPROSLA\\_v3.pdf](https://slls.eu/wp-content/uploads/2012/12/Ethics-issues-in-sign-language-acquisition-data-archiving-IPROSLA_v3.pdf)
- Branchini, C., and Donati, C. (2016). Assessing lexicalism through bimodal eyes. *Glossa* 1, 1–30. doi: 10.5334/gjgl.29
- Campbell, R., MacSweeney, M., and Waters, D. (2008). Sign language and the brain: a review. *J. Deaf Stud. Educ.* 13, 3–20. doi: 10.1093/deafed/enn035
- Chen Pichler, D., and Lillo-Martin, D. (2018). “Hearing parents and deaf children learning a sign language together” in *Talk at the 4th International Congress on Family-Centered Early Intervention for Children who are Deaf or Hard-of-Hearing*; Bad Ischl, Austria; June 15, 2018.
- Corina, D. P., and Knapp, H. P. (2006). “Psycholinguistic and neurolinguistic perspectives of sign language” in *Handbook of psycholinguistics*. 2nd Edn, eds. M. J. Traxler and M. A. Gernsbacher (Amsterdam: Elsevier), 1001–1024.
- Corina, D. P., and Spotswood, N. (2012). “Neurolinguistics” in *Sign language. An international handbook*. eds. R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 739–762.
- Cormier, K., Schembri, A., Vinson, D., and Orfanidou, E. (2012). First language acquisition differs from second language acquisition in prelingually deaf signers: evidence from sensitivity to grammaticality judgements in British sign language. *Cognition* 124, 50–65. doi: 10.1016/j.cognition.2012.04.003
- Cormier, K., Smith, S., and Sevcikova, Z. (2015). Rethinking constructed action. *Sign Lang. Linguist.* 18, 167–204. doi: 10.1075/sll.18.2.01cor
- Costello, B., Fernández, J., and Landa, A. (2008). “The non-(ex)istent native signer: sign language research in a small population” in *Sign Languages: Spinning and Unravelling the Past, Present and Future. TISLR 9: Papers from the 9th Theoretical Issues in Sign Language Research Conference, Florianópolis, Brazil, December 2006*. ed. R. Quadros (Petrópolis/RJ, Brasil: Editora Arara Azul), 77–94.
- Crasborn, O. (2015). “Transcription and notation methods” in *Research methods in sign language studies*. eds. E. Orfanidou, B. Woll, and G. Morgan (Malden, MA: Wiley Blackwell), 74–88.
- Dye, M. W. G. (2012). “Processing” in *Sign language. An international handbook*. eds. R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 687–711.
- Emmorey, K. (1999). “Do signers gesture?” in *Gesture, speech, and sign*. eds. L. S. Messing and R. Campbell (Oxford: Oxford University Press), 133–159.
- Emmorey, K. (2002). *Language, cognition, and the brain: Insights from sign language research*. Mahwah, NJ: Erlbaum.
- Emmorey, K. (2003). “The neural systems underlying sign language” in *Deaf studies, language, and education*. eds. M. Marschark and P. E. Spencer (Oxford: Oxford University Press), 361–376.
- Emmorey, K., and Corina, D. (1990). Lexical recognition in sign language: effects of phonetic structure and phonology. *Percept. Mot. Skills* 71, 1227–1252.
- Emmorey, K., Giezen, M. R., and Gollan, T. H. (2016). Psycholinguistic, cognitive, and neural implications of bimodal bilingualism. *Biling. Lang. Cogn.* 19, 223–242. doi: 10.1017/s1366728915000085
- Engberg-Pedersen, E. (1993). *Space in Danish sign language: The semantics and morphosyntax of the use of space in a visual language*. Hamburg: Signum.
- Emmorey, K., Thompson, R., and Colvin, R. (2009). Eye gaze during comprehension of American sign language by native and beginning signers. *J. Deaf Stud. Deaf. Educ.* 14, 237–243. doi: 10.1093/deafed/enn037
- Freel, B. L., Clark, M. D., Anderson, M. L., Gilbert, G. L., Musyoka, M., and Hauser, P. (2011). Deaf individuals’ bilingual abilities: American sign language proficiency, reading skills, and family characteristics. *Psychology* 2, 18–23. doi: 10.4236/psych.2011.21003
- Frishberg, N., Hoiting, N., and Slobin, D. I. (2012). “Transcription” in *Sign language. An international handbook*. eds. R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 1045–1075.
- Geraci, C., Bayley, R., Cardinaletti, A., Cecchetto, C., and Donati, C. (2015). Variation in Italian sign language (LIS): the case of wh-signs. *Linguistics* 53, 125–151. doi: 10.1515/ling-2014-0031
- Goldin-Meadow, S., and Brentari, D. (2017). Gesture, sign and language: the coming of age of sign language and gesture studies. *Brain Behav. Sci.* 39, 1–17. doi: 10.1017/s0140525x15001247
- Grosvald, M., Gutierrez, E., Hafer, S., and Corina, D. (2012). Dissociating linguistic and non-linguistic gesture processing: electrophysiological evidence from American sign language. *Brain Lang.* 121, 12–24. doi: 10.1016/j.bandl.2012.01.005
- Hänel-Faulhaber, B., Kügow, M., Skotara, N., Salden, U., and Röder, B. (2014). ERP correlates of German sign language processing provide evidence for sensitive/critical periods of language learning. *BMC Neurosci.* 15:62. doi: 10.1186/1471-2202-15-62
- Hanke, T. (2016). “Towards a visual sign language corpus linguistics” in *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining, Language Resources and Evaluation Conference (LREC)*, 89–92.
- Hanke, T., Matthes, S., Regen, A., and Wörseck, S. (2012). “Where does a sign start and end? Segmentation of continuous signing” in *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interaction between Corpus and Lexicon, Language Resources and Evaluation Conference (LREC)*, 69–74.
- Hosemann, J. (2011). Eye gaze and verb agreement in German sign language: a first glance. *Sign Lang. Linguist.* 14, 76–93. doi: 10.1075/sll.14.1.05hos
- Hosemann, J. (2015). The processing of German sign language sentences. Event-related potential studies on phonological, morpho-syntactic, and semantic aspects. Ph.D. dissertation. Universität Göttingen.
- Hosemann, J., Herrmann, A., Steinbach, M., Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2013). Lexical prediction via forward models: N400 evidence from German sign language. *Neuropsychologia* 51, 2224–2237. doi: 10.1016/j.neuropsychologia.2013.07.013
- Hosemann, J., Herrmann, A., Sennhenn-Reulen, H., Schlesewsky, M., and Steinbach, M. (2018). Agreement or no agreement. ERP correlates of verb agreement violations in German sign language. *Lang. Cognit. Neurosci.* doi: 10.1080/23273798.2018.1465986
- Johnston, T. (2006). W(h)ither the deaf community? Population, genetics, and the future of Australian sign language. *Sign Lang. Stud.* 6, 137–173. doi: 10.1353/sls.2006.0006
- Labov, W. (1996). “When intuitions fail” in *Papers From the Parasession on Theory and Data in Linguistics Chicago Linguistic Society, Vol. 32*. eds. L. McNair, K. Singer, L. Dolbrin, and M. Aucon (Chicago, IL: Chicago Linguistic Society), 77–106.
- Liddell, S. K. (2003). *Grammar, gesture, and meaning in American sign language*. Cambridge: Cambridge University Press.
- Liddell, S. K., and Metzger, M. (1998). Gesture in sign language discourse. *J. Pragmat.* 30, 657–697. doi: 10.1016/S0378-2166(98)00061-7
- Lieberman, A. M., Borovsky, A., Hatrak, M., and Mayberry, R. I. (2015). Real-time processing of ASL signs: delayed first language acquisition affects organization of the mental lexicon. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 1130–1139. doi: 10.1037/xlm0000088
- Lieberman, A. M., Borovsky, A., and Mayberry, I. R. (2017). Prediction in a visual language: real-time sentence processing in American sign language across development. *Lang. Cognit. Neurosci.* 1–15. doi: 10.1080/23273798.2017.1411961
- Lillo-Martin, D. (2018). Differences and similarities between late first-language and second-language learning. *Biling. Lang. Cognit.* 21, 924–925. doi: 10.1017/s1366728918000159
- Lillo-Martin, D., Müller de Quadros, R., and Chen Pichler, D. (2016). The development of bimodal bilingualism: implications for linguistic theory. *Linguist. Approaches Biling.* 6, 719–755. doi: 10.1075/lab.6.6.01lil
- Macht, C., and Steinbach, M. (2018). “Regionalsprachliche merkmale in der deutschen gebärdensprache” in *Sprache und raum. Ein internationales handbuch der sprachvariation*. eds. J. Herrgen and J. E. Schmidt (Berlin: De Gruyter Mouton).
- Mathur, G., and Rathmann, C. (2006). “Variability in verbal agreement forms across four signed languages” in *Laboratory phonology VIII*. eds. L. Goldstein, D. Whalen, and C. T. Best (Berlin: Mouton), 287–314.
- Matthewson, L. (2004). On the methodology of semantic fieldwork. *Int. J. Am. Linguist.* 70, 369–415. doi: 10.1086/429207

- Mayberry, R. I. (2010). "Early language acquisition and adult language ability: what sign language reveals about the critical period for language" in *Oxford handbook of deaf studies, language, and education*, Vol. 2. eds. M. Marschark and P. Spencer (Oxford: Oxford University Press), 281–291.
- Mayberry, R. I., and Kluender, R. (2017). Rethinking the critical period for language: new insights into an old question from American sign language. *Biling. Lang. Cognit.* 21, 886–905. doi: 10.1017/s1366728917000724
- Meier, R. P. (2002). "Why different, why the same? Explaining effects and non-effects of modality upon linguistic structure in sign and speech" in *Modality and structure*. eds. R. P. Meier, et al. (Cambridge: Cambridge University Press), 1–25.
- Meier, R. P. (2012). "Language and modality" in *Sign language. An international handbook*. eds. R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 574–601.
- Mitchell, R. E., and Karchmer, M. A. (2004). Chasing the mythical ten percent: parental hearing status of deaf and hard-of-hearing students in the United States. *Sign Lang. Stud.* 4, 138–163. doi: 10.1353/sls.2004.0005
- Muir, L. J., and Richardson, I. E. (2005). Perception of sign language and its application to visual communications for deaf people. *J. Deaf. Stud. Deaf. Educ.* 10, 390–401. doi: 10.1093/deafed/eni037
- Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., and Lee, R. G. (2000). *The syntax of american sign language. Functional categories and hierarchical structure*. Cambridge, MA: MIT Press.
- Nishio, R., Hong, S.-E., König, S., Konrad, R., Langer, G., Hanke, T., et al. (2010). "Elicitation methods in the DGS (German sign language)" in *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, 178–185.
- Orfanidou, E., Woll, B., and Morgan, G. (2015). *Research methods in sign language studies*. Malden, MA: Wiley Blackwell.
- Padden, C. (1998). *Interaction of morphology and syntax in American sign language*. New York: Garland.
- Pfau, R., and Steinbach, M. (2011). "Grammaticalization in sign languages" in *Handbook of grammaticalization*. eds. B. Heine and H. Narrog (Oxford: Oxford University Press), 681–693.
- Pfau, R., and Steinbach, M. (2016). Meaning and modality. Plurality of relations in German sign language. *Lingua* 170, 69–91. doi: 10.1016/j.lingua.2015.11.002
- Poizner, H., Klima, E., and Bellugi, U. (1987). *What the hands reveal about the brain*. Cambridge, MA: MIT Press.
- Quadros, R. M. D. (2018). Bimodal bilingual heritage. signers: a balancing act of languages and modalities. *Sign Lang. Stud.* 18, 355–384. doi: 10.1353/sls.2018.0007
- Quer, J., Cecchetto, C., Pfau, R., Donati, C., Steinbach, M., Geraci, C., et al. (2017). *SignGram blueprint. A guide to sign language grammar writing*. Berlin: De Gruyter Mouton Available (Open Access) at <https://www.degruyter.com/viewbooktoc/product/467598?rskey=kMkMY6&result=1>
- Sáfar, E., and Glauert, J. (2012). "Computer modelling" in *Sign language. An international handbook*. eds. R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 1075–1101.
- Sandler, W., and Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge: Cambridge University Press.
- Schembri, A., and Lucas, C. (2015). *Sociolinguistics and deaf communities*. Cambridge: Cambridge University Press.
- Schermer, T. (2012). "Language planing" in *Sign language. An international handbook*. eds. R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 889–908.
- Schlenker, P. (2018). Visible meaning: sign language and the foundations of semantics. *Theor. Linguist.* 44, 123–208. doi: 10.1515/tl-2018-0012
- Skotara, N., Salden, U., Kügow, M., Hänel-Faulhaber, B., and Röder, B. (2012). ERP correlates of early language deprivation: a comparison of native signers and signers with delayed language acquisition in processing German as L2. *BMC Neurosci.* 13:44. doi: 10.1186/1471-2202-13-44
- Steinbach, M., and Onea, E. (2016). A DRT analysis of discourse referents and anaphora resolution in sign language. *J. Semant.* 33, 409–448. doi: 10.1093/jos/ffv002
- Taub, S. F. (2012). "Iconicity and metaphor" in *Sign language. An international handbook*. eds. R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 388–412.
- Thompson, R., Emmorey, K., and Kluender, R. (2006). The relationship between eye gaze and verb agreement in American sign language: an eye-tracking study. *Nat. Lang. Linguist. Theory* 24, 571–604. doi: 10.1007/s11049-005-1829-y
- Thompson, R., Emmorey, K., and Kluender, R. (2009). Learning to look: the acquisition of eye gaze agreement during the production of ASL verbs. *Biling. Lang. Cogn.* 12, 393–409. doi: 10.1017/s1366728909990277
- Thompson, R., Vinson, D., Fox, N., and Vigliocco, G. (2013). "Is lexical access driven by temporal order or perceptual salience? Evidence from British sign language" in *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*. Vol. 35, No. 35, 1450–1455.
- van Herreweghe, M., and Vermeerbergen, M. (2012). "Handling sign language data" in *Sign language. An international handbook*. eds. R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 1023–1045.
- Wienholz, A., Nuhbalaoglu, D., Mani, N., Herrmann, A., Onea, E., and Steinbach, M. (2018). Pointing to the right side? An ERP study on anaphora resolution in German sign language. *PLoS ONE* 13:e0204223. doi: 10.1371/journal.pone.0204223
- Woll, B. (2012). "Atypical signing" in *Sign language. An international handbook*. eds. R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 762–787.
- Woll, B. (2018). The consequences of very late exposure to BSL as an L1. *Biling. Lang. Cognit.* 21, 936–937. doi: 10.1017/s1366728918000238

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Quer and Steinbach. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

Visit us: [www.frontiersin.org](http://www.frontiersin.org)

Contact us: [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

[@frontiersin](https://www.instagram.com/frontiersin)



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership