Typical and Impaired Processing in Morphosyntax

Edited by
Vincent Torrens

Language Acquisition & Language Disorders
64

John Benjamins Publishing Company

Typical and Impaired Processing in Morphosyntax

Language Acquisition and Language Disorders (LALD)

ISSN 0925-0123

Volumes in this series provide a forum for research contributing to theories of language acquisition (first and second, child and adult), language learnability, language attrition and language disorders.

For an overview of all books published in this series, please see *benjamins.com/catalog/lald*

Volume 64

Typical and Impaired Processing in Morphosyntax
Edited by Vincent Torrens

Typical and Impaired Processing in Morphosyntax

Edited by

Vincent Torrens
National University of Distance Learning

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/lald.64

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2020027504 (PRINT) / 2020027505 (E-BOOK)

ISBN 978 90 272 0763 0 (HB)
ISBN 978 90 272 6066 6 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of contents

Determiner Phrases

Language impairment

Introduction

Typical and impaired processing of relative clauses, empty categories and determiner phrases

Vincent Torrens
Universidad Nacional de Educación a Distancia

Language processing is considered to be an important part of cognition and studies conducted in this field are increasing. The present volume deals with research on language processing and disorders. It covers topics ranging across syntax processing, bilingualism, lexical processing and language disorders. The articles presented in this volume cover a number of linguistic phenomena, including the following: subject control, object control, raising, unaccusatives and unergatives, noun gender, relative clauses, A′ movement, filler-gap dependencies, ditransitives, prepositional phrases, argument structure, personal pronouns, anaphora, long distance extraction and inflectional morphology.

The research presented in this volume covers typical language processing, child developmental language disorders, adult neurodegenerative disorders and neurological bases of typical or impaired brains. Syntax processing has been described by processing models such as the garden path model, constraint-based models, the good-enough theory, serial and parallel processing, modular and interactive theories of language processing. The garden path model (Frazier, 1987) proposes that a single parse is constructed in a serial manner. When an initial parse of a sentence turns out to be incorrect due to syntactic ambiguity, a reanalysis of the syntactic parse takes place. During reanalysis, some principles take effect, such as late closure or minimal attachment; late closure states that words are attached to the current clause being processed; minimal attachment states that the parser builds the simplest syntactic structure possible in ambiguous sentences. Constraint-based theories are based on statistical learning: the parser takes into account the frequencies and distribution of linguistic structures; speakers apply probabilistic constraints when faced with ambiguous sentences. The good-enough models propose that listeners apply partial and superficial representations and that they do not apply a detailed syntactic analysis to sentences.

https://doi.org/10.1075/lald.64.int

Papers in this collection use a variety of experimental methods, such as eye tracking, reaction times, Event Related Potentials, picture selection tasks, sentence elicitation and picture matching tasks. Eye tracking is a method which measures where a participant looks, giving information about the point of gaze. The eye tracker measures the position of the eyes and their movements. Infrared light is directed to the pupils of the eyes, and the reflections are tracked by a camera. Reaction time is a measure of the speed with which a participant reacts to an item: when a participant has to respond to a more difficult sentence, the reaction time is typically longer compared to an easier sentence. Event-related potentials (ERPs) measure the electrical activity in the scalp in a reaction to an event. The event is a stimulus like a sound, a word or a sentence, sometimes followed by a question where the participants need to select an answer or remember an item. The studies included in this book try to cover some of the most representative methods used in language processing.

The papers of this volume are organized into four groups: (i) the processing of relative clauses; (ii) empty categories; (iii) determiner phrases, and (iv) language impairment. This last group considers some of the same syntactic phenomena but from the perspective of individuals with language impairment.

The first paper in the section on relative clauses is the study by **Jana Mewe**, which addresses the question of how free relatives are interpreted by adults speakers of German. Free relative clauses do not have a head. Rather the whole free relative clause functions as a complement of the transitive verb in object position. Constructions in which the matrix verb governs a less marked case than the verb of the free relative clause (hh for hierarchy harmonic) are judged as more acceptable than constructions in which it is the other way around (ho for hierarchy opposing); that is, constructions are more difficult to process when the cases of the covert head and the relative pronoun differ. However, case-matching conditions are, unsurprisingly, always judged more acceptable than any mismatches; that is, constructions where the case of the covert head and the relative pronoun is the same are easier to process. The experiment examines both case-mismatching and case-matching, involving accusative and dative case. The study applies the methodology of self-paced reading in incremental processing. The authors found that case-matching for datives, unexpectedly, was read significantly faster than other conditions, including ho-case-mismatches. The author concludes that there must be other mechanisms at work that apply to dative case; markedness and idiosyncracy provide an advantage for dative case-matching.

Another paper on the processing of relative clauses is the work by **Anastasia Paspali**, which explores the psychological reality of traces. This author considers two processing accounts to describe filler integration: the trace reactivation hypothesis (TRH) and the direct association hypothesis (DAH). According to the

TRH, filler integration is modulated by syntactic traces, whereas DAH proposes a link between the filler and its verb, which is driven by semantic information. In this study, the paradigms Cross-modal Priming (CMP) and Probe Classification During Reading (PCDR) are applied. Paspali tested ditransitive constructions where the indirect object (IO) surfaces as a prepositional phrase in Greek. The results confirm that filler integration is modulated by syntactic traces, that the IO–DO order is base generated (i.e. IOs are higher hierarchically than DOs) and that the CMP paradigm is sensitive to test filler integration.

The last study on the processing of relative clauses is the paper by **Sandra Villata** and **Paolo Lorusso**, which deals with how relative clauses are interpreted in Italian. The difficulty of interpretation of Italian relative clauses differs between a subject and an object reading, depending on many factors: since longer dependencies have a greater processing cost than shorter ones, the parser first engages in a subject analysis, but the object interpretation is attainable through reanalysis. Cues to reanalysis that occur late in the sentence are expected to be less efficient as compared to early cues, under self-organized parsing accounts, while no effect of cue timing is predicted under standard syntactic and processing accounts. The authors tested the effectiveness of a late cue in triggering a reanalysis by testing gender, which can be manipulated in the embedded clause. A self-paced reading experiment, combined with a sentence comprehension task, was conducted with Italian native speakers. Gender agreement on the embedded past-participle was manipulated, in order to elicit either a subject reading or an object reading. Results show that participants were both slower and less accurate in the object reading condition. Their results suggest that once that the parser stabilizes on the default subject reading, later cues are ineffective in triggering an object reanalysis.

The second section, on empty categories, explores the possibility that experimental evidence can provide support for the existence of these categories. Basically, these studies explore reactivation of traces, as well as priming effects. Previous studies have suggested that processing is by the reactivation of a trace (e.g., Nicol & Swinney, 1989). Sentences containing moved constituents, such as wh-questions, are usually referred to as involving filler-gap dependencies. The processor tries to link a filler (such as a wh-phrase) to a possible gap in the sentence (the position from which it has moved). The dependency between the filler and the gap can be semantic or syntactic. In previous studies, it has been found that different Event Related Potential components are related to different linguistic properties. One of these components is the N400, which is found when a semantic anomaly is detected; on the other hand, when a syntactic anomaly or a syntactically complex sentence is detected, it causes a P600 component. **Felser and Jessen** ran two experiments in order to investigate direct object filler-gap dependencies in German. The expectation was that sentences containing a semantically implausible object filler would elicit an N400 response at the filler's potential lexical-semantic licenser, and sentences containing a filled object gap would elicit a P600 component. Felser and Jessen tested 21 German native speakers. The experimental items contained a main clause headed by a transitive verb, followed by an object relative clause introduced by a relative pronoun. The filler NP was either a plausible or an implausible direct object of the relative clause. The brain activity to the critical verb produced an N400 component for implausible over plausible sentences in the implausible filler condition, whereas a P600 was produced in the filled gap condition. These results suggest that the semantic and syntactic integration processes involved in establishing filler-gap dependencies are separable and independent.

In **Tori Larsen** and **Christer Johansson**'s paper, an account of the status of PRO is provided. PRO is an empty category that fills the subject position in non-finite clauses. These authors propose that this position is filled by a trace instead of the empty category PRO. They present findings on possible reactivation patterns in control sentences, using picture recognition of PRO antecedents presented either before or after the infinitive marker. One image is presented in one of the two positions and participants are asked to recognize if the written equivalent of the image was present previously in the sentence or not. The pictures that are previously presented in the same sentence can be a correct antecedent of either PRO or the trace of a raising sentence. The results support the proposal that raising sentences and control sentences behave similarly under syntactically primed conditions. This research gives support to a movement account of both control and raising.

The goal of **Joakim Vea** and **Christer Johansson**'s paper is to present new experimental findings regarding neuter forms of adjectives in the Mainland Scandinavian languages. These forms are formed by a simple rule: add /t/. This rule can be applied almost without exception; however, a small group of adjectives appear to have no well-formed neuter. The experiment was conducted on a sample of 30 native speakers of Norwegian. Using a masked priming paradigm, these authors tested the priming effect of expected neuter forms on stem forms of the same lexeme, in a lexical decision task. Within the same task, the priming effect for missing neuter forms was compared to that of adjectives with regular neuters, as well as to pseudowords with constructed regular neuters. Adjectives were analyzed as a condition with three levels: normal, problematic (i.e., without neuter) and non-words (i.e., pseudowords). Priming effects were apparent for each group, with all groups showing different levels of priming; the strongest facilitation occurred in the normal group. The authors propose that the priming stimulus activates a lexical or morphological process for lexical items only, while priming nonwords with a nonword facilitates a decision for nonwords.

The section on determiners includes two papers on the processing of case checking and noun gender. The paper by **Loes Koring**, investigates how adults and children check nominative and accusative case in unaccusative and unergatives sentences in English and Dutch. The class of intransitive verbs can be partitioned into verbs that take an internal argument (unaccusatives, like "fall") and verbs that do not (unergatives, like "jump"). One difference between unaccusative verbs and unergative verbs is in the role they assign to their argument. Whereas "fall" assigns the role of theme to its argument, "jump" assigns the role of agent. Even though an agentive reading for "fall" is possible, this reading is in fact harder for listeners to get than a non-agentive reading, since it requires coercion. A puppet presented sentence pairs to the 4-year-old participants, while adults received a written version. The task was to judge which of the sentences sounded silly. Both adults and children judged the sentences that require an agentive NP argument for an unaccusative verb indeed to be silly. For adults this was true in 100% of the trials, for children on 74% of the trials. This is hard to reconcile with a parser that is guided by a heuristic which dictates that the first NP is an agent. It is far from clear that extra-grammatical processing heuristics reduce processing cost, or that our sentence representations are based on them.

The paper by **Elena Tribushinina, Julia Lomako, Natalia Gagarina, Ekaterina Abrosova** and **Pim Mak** investigates the difficulty simultaneous bilinguals have in acquiring grammatical gender, especially in their weaker (non-dominant) language. Specifically, these authors studied the production of adjective-noun agreement. They studied the online processing of gender cues by simultaneous Dutch-Russian bilinguals raised in the Netherlands. This language combination is theoretically interesting, since Russian has a three-way system of grammatical gender determining pronoun use, whereas, in present-day Dutch, pronoun selection is primarily based on semantic cues, such as animacy and individuation. In order to establish whether bilingual children can use gender cues in online processing of pronouns (in their weaker language, Russian), an eye-tracking experiment was conducted by means of the Visual World Paradigm. The participants saw two pictures (e.g., a monkey and an ant) and heard sentences such as 'The ant sees the monkey. SHE wants a new toy.' The nouns in the first clause were either of the same or of different genders. In the same-gender condition, the pronoun was ambiguous. Only in the different gender condition was the pronoun informative (i.e., referring either to the subject or to the object of the first clause). The authors observed that in the same-gender (ambiguous) condition the monolingual participants tended to look at the subject of the previous clause, whereas the bilingual group did not have a preference. In the different-gender condition bilingual children, just like their monolingual peers, were sensitive to gender cues. However, in the bilingual group the effect was observed later than in the monolingual group, which might reflect an additional processing effort, possibly related to the need to suppress the dominant language.

The section on language impairment includes papers on developmental and neurodegenerative disorders. Language impairment can in general be classified as involving acquired, developmental or neurodegenerative disorders. Acquired language disorders, such as those due to stroke or traumatic brain injury, are not covered in this section. Developmental disorders include Specific Language Impairment, Autism, Down Syndrome, Hearing Impairment, Williams Syndrome, X fragile Syndrome, Prader-Willis, Syndrome, Angelman Syndrome or Developmental Dyslexia (Torrens, 2018). The category of neurodegenerative disorders includes diseases such as Alzheimers, Huntingtons or Parkinsons. An accurate analysis of the language properties in the syndromes described in these chapters will allow us to arrive at an accurate interpretation of symptoms. The first paper in this section is the study by **Silvia d'Ortenzio, Silvia Montino, Alessandro Martini, Patrizia Trevisi** and **Francesca Volpato**, which consists of research on three Italian-speaking children suffering from severe or profound sensorineural hearing loss. The three children have cochlear implants. Relative clauses are known to be difficult for children with hearing loss, even with hearing aids or choclear Implants. Previous studies on aphasic patients and children with specific language impairment have shown improvement of complex structures after explicit teaching of syntactic rules. This study investigates the production of subject and object relative clauses in Italian, by means of three case studies, in a longitudinal design. The authors apply a teaching intervention, checking whether there is improvement in proficiency in relative clauses and narrative skills. Results are compared with typically developing children. The authors use a preference task, an elicitation task and storytelling. The production and comprehension of relative clauses improved after the intervention and the authors found generalization effects to untrained structures.

Natascha Lantschner and **Anna Cardinaletti** study an Italian girl with a sex chromosome anomaly, Trisomy X, diagnosed with Expressive Language Impairment. According to these authors, the participant had difficulties with function words, such as articles, prepositions, and clitic pronouns, while at the same time showing has good cognitive abilities. Oral comprehension, oral production, repetition, grammaticality judgement, working memory and auditory discrimination were assessed. After an initial evaluation, the authors applied a linguistic intervention, in order to stimulate language awareness and to promote the acquisition of articles and prepositions through explicit teaching. A second assessment of the child's language competence was carried out after the intervention: they found that the comprehension of narratives, passive clauses, relatives sentences and cleft sentences improved; with respect to production and repetition, the production of clitic pronouns did not improve, the accuracy in Wh-questions slightly increased; in the repetition task, the child produced more articles, passive sentences and complementizers; the production of articles and prepositions increased; however, the performance on grammaticality judgements, working memory and auditory discrimination did not improve.

The contribution by **María Teresa Martín-Aragoneses, David del Río, Ramón López-Higes, José María Prados, Pedro Montejo** and **María Luisa Delgado-Losada**, consists of a study of sentence processing in individuals with Mild Cognitive Impairment. Mild Cognitive Impairment is a disorder that consists of a slight decline in cognitive abilities, including memory, language and thinking skills. The methodology used in this study consisted of a self-paced reading task and neuropsychological measures, including Working Memory and Interference Control. These authors measure reading times and sentence comprehension in young adults, cognitively intact older adults and older adults with Mild Cognitive Impairment. Syntactic complexity was examined by using embedded relative clauses varying in two features: (a) subject extraction versus object extraction of the antecedent noun-phrase; and (b) short-distance versus long-distance dependencies between the extracted object and the embedded verb. The authors conclude that Working Memory capacity explains variability among groups for comprehension. In conclusion, results suggest that Interference Control and Working Memory play a role in the processing and comprehension of syntactically complex sentences in older adults.

The aim of the final study, by **Eva Wimmer** and **Martina Penke**, was to conduct experimental research on the acquisition of wh-questions and passives in 22 German-speaking children and adolescents with Down Syndrome (DS). In particular, they conducted a standardized passive comprehension test and a non-standardized comprehension test on wh-questions. Only eight children with DS performed above chance on passive clauses as well as on wh-questions while the other 14 children with DS had severe comprehension difficulties related to at least one of these sentence types. Passive sentences were more affected than questions. The results suggest that an impairment in the syntactic development of children with DS affects also A' movement constructions (as wh-questions), not only structures involving A movement chains (as passives). The impairment occurs in a substantial proportion of children with DS and cannot simply be attributed to their general cognitive disabilities.

In conclusion, the present volume presents research on language processing that was presented at the Experimental Psycholinguistics Conference in Menorca (EPC) in June 2017 and at The Conference on Developmental Language Disorders in Madrid (DevO) in September 2018. I would like to thank the plenary speakers of both Conferences (Stephanie Durrleman, Paul Fletcher, Maria Teresa Guasti, Cornelia Hamann, Luigi Rizzi and Lydia White) and the members of the scientific and organizing committees (Eva Aguilar, Lluís Barceló-Coblijn, Mercedes

Belinchón, Adriana Belletti, Anna Cardinaletti, Tatiana Chernigovskaya, Vicky Chondrogianni, Iris Duinmejer, Carlos Gallego, John Grinstead, Kleanthes Grohmann, Pedro Guijarro, Cornelia Hamann, Kristine Jensen de López, Melita Kovacevic, Paulina Łęska, Mayte Martín-Aragoneses, Maria Mastropavlou, Martina Penke, Alexandra Perovic, Mihaela Pirvulescu, Philippe Prévost, Esther Ruigendijk, David Saldaña, Jeannette Schaeffer, Carson Schütze, Arhonto Terzi, Vincent Torrens and Laurie Tuller). I'd like to thank all the anonymous reviewers of this volume and also the contribution of the coordinators of this book series Roumyana Slabakova and Lydia White. I am also very grateful to Universidad Nacional de Educación a Distancia and to the Cultural Council of Menorca for their support.

References

Bever, T. G., & McElree, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry*, 19, 35–43.

Cuetos, F., & Mitchell, D. C. (1988). Crosslinguistic differences in parsing. *Cognition*, 30, 73–105. https://doi.org/10.1016/0010-0277(88)90004-2

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51, 97–114. https://doi.org/10.1016/j.jml.2004.02.003

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18(1):5–19. https://doi.org/10.1007/BF01069043

Torrens, V. (Ed.). (2018). Identification of children with Specific Language Impairment. *Language Acquisition*, 25(1), 1–117. https://doi.org/10.1080/10489223.2016.1191078

Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/6184.001.0001

Relative clauses

Case(mis)matching in German free relative clauses in the self-paced reading paradigm

Jana Mewe
University of Cologne

Off-line measures suggest that the *hierarchy rule strategy* applies to the processing of free relative clauses (RCs) in German, suggesting that case matches between covert head and RC-pronoun are preferred, whereas mismatches are dissociated based on their (non)conformity with the case hierarchy. We asked whether on-line sentence processing supports this pattern, and investigated free RCs with combinations of nominative, accusative and dative case in the self-paced reading paradigm. Our results are captured best by assuming two further strategies: the *parallel syntactic function strategy* and, especially, the *subject-first strategy.* Based on these strategies, we provide an Optimality Theory-style analysis that explains prior findings and particularities of incremental sentence processing, and that additionally accounts for the exceptional behaviour of accusative-nominative mismatches.

Keywords: German free relative clauses, headed relative clauses, case, incremental processing, self-paced reading, thematic roles, parallel syntactic function strategy, subject first strategy, case hierarchy

1. Introduction

A number of forms of nominal modification are attested across languages, including relative clauses. German allows for headed relative clauses, which have an overt head, and free relative clauses, which do not have an overt head, as shown in (1). Free relatives are the focus of the current paper.

(1) *Der Vater umarmte ___, wen der Sohn korrigiert hatte.*
the father hugged ___$_{acc}$ whom$_{acc}$ the$_{nom}$ son$_{nom}$ corrected had
'The father hugged whoever the son had corrected.'

https://doi.org/10.1075/lald.64.01mew

Free relative clauses have been an issue in theoretical and empirical research for decades. However, while previous empirical research made use of off-line methods (acceptability ratings and corpus studies) the current paper will pay attention to the influence of incremental processing of German free relative clauses by making use of self-paced reading. Another hitherto neglected aspect is that previous research on free relative clauses in German has rarely considered commonalities they share with relative clauses with overt heads. One conspicuous trait they share is the preference for a parallel syntactic function of nominal head and relative pronoun that has been observed in typological research for relative clauses with overt heads (Kirby, 1998; Newmeyer, 2005). The preference for case match in free relative clauses can be seen as a special case-based manifestation of this typological preference. This typological preference has been shown to be operative in German relative clauses with overt heads. Fanselow et al. (1999), for example, summarise a reading time experiment revealing that case-ambiguous grammatical functions were preferentially assigned to the relative pronoun on the basis of the case borne by the head noun in German.

This preference in relative clauses with overt heads was previously thought to be more rigorously applied in free relative clauses. Gross and van Riemsdijk (1981; see also van Riemsdijk 2006) assumed German free relative clauses to be grammatical only if they display parallel syntactic functions (as in (1)), that is, where the case of the covert head and the relative pronoun are identical.

In (1) the head of the relative clause is overtly missing (marked by ___); it is covert, hence, the name *free* (or *headless*) *relative clause*. The object position of the verb of the matrix clause is not filled and, thus, lacks an overt realisation of its case. However, the covert case in the matrix clause can be reconstructed by the case of the relative pronoun. This is accomplished most easily when the two cases match, as in (1).

Bausewein's (Bausewein 1991)[1] findings contradict the initial claim that only free relative clauses such as given in (1) are grammatical. Thus, the phenomenon is more complex than originally assumed. Consider the examples for case mismatches in (2):

(2) Case mismatches in German
a. *Keine Geduld besaß* ___, *wen letztens der Sohn*
No patience had ___$_{nom}$, whom$_{acc}$ recently the$_{nom}$ son$_{nom}$
korrigiert hatte.
corrected has.
'No patience had, whom the son has recently corrected.'

1. Bausewein later changed her name to Pittner.

b. *Der Vater umarmte___, wer letztens den Sohn*
The father hugged $\_\__{acc}$, who$_{nom}$ recently the$_{acc}$
korrigiert hatte.
son$_{acc}$ corrected has.
'The father hugged, who the son has recently corrected. '

Derived from her findings, Bausewein (Bausewein 1991) established a *hierarchy rule* (see (3)) that is based on the case hierarchy given in (4) to account for deviations from free relative clauses featuring case matches:

(3) Hierarchy rule:
In a case conflict between the case governed by the matrix verb (in the following: *case-matrix*) and the case governed by the verb of the free relative clause (in the following: *case-relative), case-matrix* can be unrealised if it precedes *case-relative* on the following hierarchy:

(4) Case hierarchy: nominative > accusative > dative[2]

According to the hierarchy rule (3), a sentence like the one given in (2a) is acceptable because the case of the covert head precedes the case of the relative pronoun on the case hierarchy. A sentence like the one given in (2b), however, does not observe this rule and, therefore, is predicted to be unacceptable by (3), contrary to fact. Constructions like (2a), which are harmonic with the hierarchy rule, will henceforth be called (*hierarchy) harmonic mismatch* and constructions like (2b), which oppose the hierarchy rule, will henceforth be called (*hierarchy) opposing mismatch*.

Pittner (2003) continued this research with a corpus study. All attested case mismatches followed the hierarchy rule except for the type illustrated in (2b): hierarchy opposing accusative-nominative mismatch (Pittner, 2003, p. 205). This corpus study points in the same direction as an earlier explorative acceptability rating by the same author (Bausewein, 1991). In sum, both studies suggest that, with a notable exception, the hierarchy rule captures the data appropriately.

The hierarchy-based approach of Pittner described above was tested in a series of experiments by Vogel and colleagues using speeded acceptability ratings (Vogel & Frisch, 2003; Vogel & Zugck, 2003; Vogel, 2011) and corpus analyses (Vogel, Frisch & Zugck, 2006). Importantly for the present study, Vogel and colleagues included comparisons of case match conditions to case mismatch conditions and reported for all experiments that case match conditions were judged significantly more acceptable than case mismatch conditions. Case match conditions were

2. Since the present paper is only concerned with nominative, accusative and dative, the other cases have been left out in (4).

never reported to differ significantly from each other. The hierarchy harmonic mismatch conditions were always judged significantly more acceptable than the hierarchy opposing mismatch conditions, including the hierarchy opposing accusative-nominative mismatch. This contradicts the findings of Pittner regarding this type of mismatch mentioned earlier.

A more recent paper-and-pencil acceptability study supports most of Vogel's (2011) findings (Mewe, 2014). Case match conditions received higher acceptability ratings than mismatch conditions and hierarchy harmonic mismatch conditions were judged significantly better than hierarchy opposing mismatch conditions. Importantly, however, the hierarchy opposing accusative-nominative mismatch illustrated in (2b) deviated from the expected pattern by not following the hierarchy rule. The comparison of the hierarchy harmonic and the hierarchy opposing mismatch for the cases nominative and accusative did not reveal a significant difference. This is remarkable because it is the same exception to the hierarchy rule Pittner (2003) found in her corpus study although Vogel's (2011) finding contradict this exception.

To sum up, previous studies provide evidence for the acceptability hierarchy in (5).

(5) match > hierarchy harmonic mismatch > hierarchy opposing mismatch

Regarding the hierarchy opposing accusative-nominative mismatch, the results are equivocal, as described above. This is noteworthy because it might hint towards a potential commonality between free and headed relative clauses: the preference for subject relative pronouns. Subject relative pronouns are preferred over object relative pronouns in relative clauses with overt NP heads in German (Schlesewsky, 1997) and universally (Keenan & Comrie, 1977 in terms of a greater accessibility of subjects over objects for relative pronouns). This strategy makes predictions that are different from the parallel syntactic function strategy, described above. While the parallel syntactic function strategy favours case match over mismatch, the accessibility strategy favours subject relative pronouns over case match and is neutral about the kind of case mismatch.

Kirby (1998) discusses accessibility and parallel syntactic function as two relative clause strategies with typological variation (see Newmeyer, 2005 for a short summary) and finds support for both (most influentially by Keenan and Comrie, 1977 for the accessibility approach and Sheldon, 1974 and MacWhinney & Pléh, 1988 for the parallel syntactic function approach). Keenan and Comrie's (1977) accessibility hierarchy favours subject relative pronouns over object relative pronouns in headed relative clauses. A problem of this approach is that the term *object* is equivocal. It may refer either to an object case, e.g. the accusative or dative, or to the semantic patient role (Primus, 2015). In terms of cases, the strategy favours only nominative relative pronouns and would disfavour accusative and dative relative pronouns in German. In terms of semantic roles, it would disfavour patient relative pronouns over agentive relative pronouns. The pertinent difference to the case constraint originates from the fact that nominative and dative may encode agentive roles in German. This is obvious for nominative arguments. However, the dative also encodes agentive roles in German (Primus, 1999; Bornkessel, 2002; Bornkessel et al., 2003). While the dative cannot encode a volitional agent; it often codes an experiencer and a possessor, which are also agentive roles due to their agentive property of sentience or possession, respectively. By contrast, the accusative is almost exclusively restricted to patient roles in modern German. Importantly, datives may occur in clause-initial basic argument position and this the obligatory position of relative pronouns in German. Datives in initial basic position are dative experiencers of psych verbs (*weil dem Kind das Geschenk gefällt* 'because the child likes the gift'), and dative possessors or experiencers with passivised verbs referring to events in which these arguments become possessors or experiencers (*weil dem Kind das Geschenk gegeben / gezeigt wurde* 'because the child was given/ shown the gift'). These assumptions are corroborated by Bornkessel's (2002) neurophysiological data that reveal a preference for datives and nominatives in initial position in German in on-line, i.e., incremental sentence processing.

As stated by the acknowledged principle of incremental processing, linguistic information, including semantic roles, is immediately processed according to the grammatical preferences of the language (e.g. Crocker, 1994; Bornkessel, 2002; Bornkessel-Schlesewsky & Schlesewsky, 2009, pp. 89–90). This means that when an accusative is encountered in initial argument position it is immediately interpreted as a patient. Due to the dispreference for patients in initial position, an accusative in this position will lead to enhanced processing costs (Bornkessel 2002; Bornkessel et al., 2003; Dröge et al., 2014)

To summarise, according to previous research the following three strategies in (6)–(8) are assumed in the present paper.

(6) The case of the covert head is higher on the case hierarchy than the case of the relative pronoun

(7) The case of the covert head matches the case of the relative pronoun

(8) Relative pronouns with a patient role in an objective case are avoided

The research presented above provides evidence that all three strategies (hierarchy rule, parallel syntactic function and no initial patients) are relevant for relative pronouns in German. However, their interaction has not been addressed. This interaction can be explored in a promising way if an analysis in Optimality Theory-style is provided. OT is a grammatical model that focusses on constraint interaction (Prince & Smolensky, 2004). Constraint interaction is captured by ranking constraints and by evaluating grammatical structures as to whether or not they comply to the highest constraint. A grammatical structure that violates the highest constraint is immediately eliminated from consideration. The evaluation procedure is documented in a tableau, which provides a powerful tool to test the relevance and ranking of specific constraints.

This is the main reason for using an OT-style analysis for our data. As we will see, the OT-analysis will not provide a perfect model of the statistical results but will come very close to them. Since the analyses refer to the conditions of the individual experiments, no candidates and no input can be provided, as regular OT requires. The conditions of the experiments are not natural candidates but the result of the experimental design. In order not to confuse the OT-style analysis with the traditional OT, slight changes concerning terminology and display will be made. That is, there will be no tableaux but tables; the pointing hand (☞) will be replaced with an arrow (→). As in traditional OT, a * indicates a constraint violation. Importantly, the OT-style analysis will provide preliminary insights into the way constraints interact with each other in incremental self-paced reading of free relative clauses in German.

2. The experiments

The current series of experiments makes use of an on-line method providing data of incremental processing: self-paced reading. The influence of incremental processing has not yet been investigated concerning free relative clauses. Self-paced reading is a method with temporal and spatial resolution in terms of where processing efforts increases during sentence comprehension. It has been repeatedly found that the time course of effects may not only increase reading times at the critical sentence region (here the relative pronoun), but may also spill over to prolong reading times at subsequent sentence regions. Based on the strategies for German free relative processing formulated in (6)–(8), three constraints and predictions derived from them are given in (9)–(11).

(9) Predictions by HIERARCHY:
Hierarchy opposing mismatches elicit longer reading times than hierarchy harmonic mismatches.

(10) Predictions by MATCH:
Case mismatches elicit longer reading times than case matches

(11) Predictions by No-PATIENT-RelPron:
Conditions with a patient role in an objective case as the relative pronoun elicit longer reading times than conditions with a thematic role having agentive features as the relative pronoun

The predictions of the individual constraints in (9)–(11) are assumed to interact with each other; however, they make contrary predictions for the conditions. While an opposing accusative-nominative mismatch is disfavoured against a harmonic nominative-accusative mismatch by MATCH (see (10)), an opposing accusative-nominative mismatch is favoured over a harmonic nominative-accusative mismatch by No-PATIENT-RelPron (see (11)). The importance of the HIERARCHY (see (9)) constraint becomes apparent in the comparison between the hierarchy harmonic nominative-dative mismatch and the hierarchy opposing dative-nominative mismatch before the final verb, disambiguating the thematic role, is encountered. MATCH and No-PATIENT-RelPron have no influence because both conditions are mismatches (that is, MATCH does not apply) and have a nominative or dative relative pronoun (that is, none of the conditions violate No-PATIENT-RelPron at this point). Thus, HIERARCHY is the only influential constraint for the comparison. While the hierarchy harmonic nominative-dative mismatch does not violate this constraint, the hierarchy opposing dative-nominative mismatch does violate it.

In the current series of experiments, we tested nominative, accusative and dative case. For practical reasons, we conducted three experiments, each focusing on a different case constellation (EXP 1: dative-accusative, EXP 2: accusative-nominative, EXP 3: dative-nominative).

Experiment 1

Experimental design and material

In order to test for case match as well as mismatch, the experiment has a 2 × 2 design, consisting of two match and two mismatch conditions. The match conditions consist of the accusative match (henceforth: mat-accacc, see (12a)) and the dative match (henceforth: mat-datdat, see (12b)). The mismatch conditions are the hierarchy harmonic mismatch (the covert head in the matrix clause is an accusative and the relative pronoun is a dative; henceforth: har-accdat, see (12c)) and the hierarchy opposing mismatch (the covert head in the matrix clause is a dative and the relative pronoun is an accusative; henceforth: opp-datacc, see 12d)).

(12) Examples for the four conditions in EXP 1 (superscripts give sentence positions for analysis)

a. accusative match (mat-accacc)

Der Vater umarmte ___,[3] *wen*^{1} *der*^{2} *Sohn*^{3} *letztens*^{4} *korrigiert*^{5} *hatte*^{6}.
The father hugged ___$_{\text{acc}}$, whom$_{\text{acc}}$ the$_{\text{nom}}$ son$_{\text{nom}}$ recently corrected has.
'The father hugged, who the son has recently corrected.'

b. dative match (mat-datdat)

Der Vater half ___, *wem*^{1} *der*^{2} *Sohn*^{3} *letztens*^{4} *vertraut*^{5} *hatte*^{6}.
The father helped ___$_{\text{dat}}$, whom$_{\text{dat}}$ the$_{\text{nom}}$ son$_{\text{nom}}$ recently trusted has.
'The father helped, whom the son has recently trusted.'

c. hierarchy harmonic accusative-dative mismatch (har-accdat)

Der Vater umarmte ___, *wem*^{1} *der*^{2} *Sohn*^{3} *letztens*^{4} *vertraut*^{5} *hatte*^{6}.
The father hugged ___$_{\text{acc}}$, whom$_{\text{dat}}$ the$_{\text{nom}}$ son$_{\text{nom}}$ recently trusted has.
'The father hugged, whom the son has recently trusted.'

d. hierarchy opposing dative-accisative mismatch (opp-datacc)

Der Vater half ___, *wen*^{1} *der*^{2} *Sohn*^{3} *letztens*^{4} *korrigiert*^{5} *hatte*^{6}.
The father helped ___$_{\text{dat}}$, whom$_{\text{acc}}$ the$_{\text{nom}}$ son$_{\text{nom}}$ recently corrected has.
'The father helped, whom the son has recently corrected.'

Each condition shown in (12a–d) has 40 lexicalisations. Due to a lack of suitable transitive verbs governing a dative, the dative verbs had to be used twice for the matrix clause and free relative clause to ensure enough experimental material. This lexical duplication was counterbalanced by grammatical and ungrammatical fillers that featured similar lexical as well as semantic duplications to distract from the test sentences. Further, garden path sentences with homonyms were used as fillers for additional distraction. No context was introduced for the test material. There were 160 test sentences (20 for each condition) distributed over eight lists, and 58 filler sentences used in each list.

3. The blank was not presented to the participants in the experiments. It is included in the examples for illustrative purposes.

Participants

96 monolingual German native speakers (12 per list) took part in the experiment after giving written consent. The study was performed in accordance with the Declaration of Helsinki and with the national and institutional recommendations adopted by the Experimental Linguistics Lab in Cologne (XLinC). The participants had normal or corrected to normal sight and were either paid for their participation or received course credit. They were naive with respect to the purpose of the experiment.

Procedure

The self-paced reading experiment was programmed and executed using OpenSesame (Mathôt et al., 2012). A non-cumulative linear word-by-word presentation with the sentence in the horizontal center was used. Every participant read 20 test sentences (5 per condition) and 58 filler sentences in pseudo-randomised order. Participants were then asked to click *yes* to continue or to answer a yes/no comprehension question (e.g. *„Did the father frighten someone?“*). Participants had five seconds to complete the task.

Results

For statistical analysis, a linear mixed-effects model with the fixed factor CASE and random intercepts for participants and items was used to analyse logarithmised reaction times (lmer(log(rt) ~ case + (1|participant) + (1|item)). Following convention, significance was set to t > 2 (Baayen, 2008). Reaction time outliers that differed more than two standard deviations from the mean were excluded from analysis. The sentence positions 1–6 are provided as indices in superscript on the German test material in Examples (12) a–d, and Figure 1 shows the corresponding mean log reaction times. Note that position 2 is a case-unambiguous nominative article (*der*), sentence position 3 is a post-case noun (see (12a–d)). We assumed that any processing cost would occur first at the relative pronoun (position 1) itself and probably at the relative clause final verb (position 5). Given the robustness of spill-over effects with self-paced reading, we also analysed the spill-over regions following the relative pronoun (positions 2 and 3) and final verb (position 6).

The only significant reading time differences were found in the spill-over regions of the relative pronoun (positions 2 and 3). At position 2, there was a significant difference between the case match conditions (mat-accacc vs. mat-datdat; $t = 3.25$), with the dative match being the condition that was read faster. The dative match was also read significantly faster than the hierarchy harmonic case mismatch (mat-datdat vs. har-accdat; $t = -3.64$), and the hierarchy opposing case mismatch (mat-datdat vs. opp-datacc; $t = 5.07$). Hence, on position 2, the dative match was read significantly faster than all other conditions which do not reveal

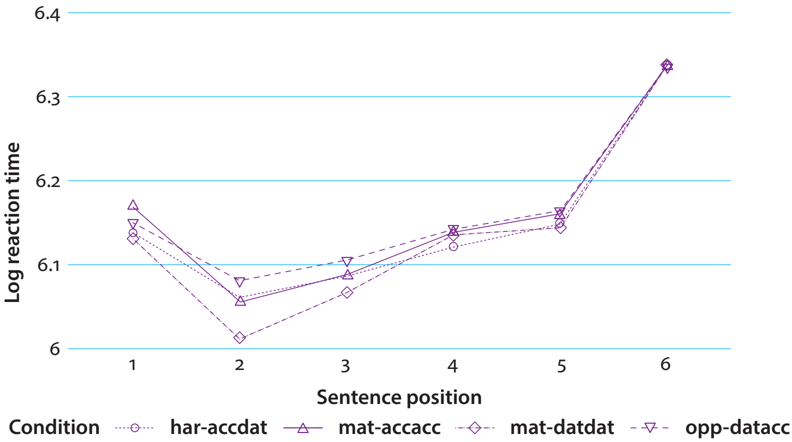

Figure 1. Logarithmised reaction time data on individually measured sentence positions

any significant reading time differences relative to each other. At position 3, the only significant reading time difference was found for the dative match being read faster than the opposing mismatch (mat-datdat vs. opp-datacc; $t = 2.56$).

Interim discussion for Experiment 1

The introduction motivated three constraints for an OT-style analysis that are assumed to interact with each other in incremental processing of free relative clauses. These three constraints are No-PATIENT-RelPron, MATCH and HIERARCHY. Note that instead of making use of the traditional OT, we only intend to follow an OT-style approach to capture the interaction of the assumed constraints. The results of position 2 (article) and 3 (noun) will be examined. First, the results of position 2 will be examined in Table 1.

Table 1. OT-style analysis for position 2

	No-PATIENT-RelPron	MATCH	HIERARCHY
mat-accacc	*		
→ mat-datdat			
har-accdat		*	
opp-datacc	*	*	*

Recall that position 2 revealed significant reading time differences of the dative match compared to all other conditions that did not differ significantly from each other. These results are accurately captured in Table 1. The dative match is the intrinsic winner of the OT-style evaluation because it does not violate any constraints. Thus, ranking is irrelevant. This is symbolised in Table 1 by missing vertical lines.

At this point, no derivations for a ranking of the constraints can be made. However, previous data from off-line studies strongly suggest the ranking MATCH >> HIERARCHY. This is consistent with the results of sentence position 2 since they do not conflict with such a ranking. The results for sentence position 3, on the other hand, necessitate a ranking of constraints; see Table 2.

Table 2. OT-style analysis for position 3

	No- PATIENT-RelPron	MATCH	HIERARCHY
→? mat-accacc	*		
→mat-datdat			
→?har-accdat		*	
opp-datacc	*	*	*

Only for the comparison of dative match (mat-datdat) and opposing mismatch (opp-datacc) was a significant reading time difference found. By tying with No-PATIENT-RelPron and MATCH, the dative match is no longer the intrinsic winner of the evaluation. This leads to two evaluations. In the first evaluation, that is No-PATIENT-RelPron >> MATCH, the dative match (mat-datdat) and the harmonic mismatch (har-accdat) would be the winners. In the second evaluation, that is MATCH >> No-PATIENT-RelPron, the match conditions (mat-accacc and mat-datdat) would be the winners. The tie results in three winners of the evaluation: accusative match (mat-accacc), dative match (mat-datdat) and harmonic mismatch (har-accdat). While this does not capture the fact that only the dative match and the opposing mismatch differed significantly from each other at position 3, it provides a very close approximation to the statistical findings. In order to signal a discrepancy between the predictions of the OT-style analysis and the statistical results, a question mark is introduced in Table 2 and subsequent tables if necessary. The hierarchy opposing mismatch (opp-datacc) is an intrinsic loser that is sorted out independently of the rank of the three constraints. However, as will become apparent from the pattern of results from EXP 2 and 3, a ranking as seen in Table 2 best captures the data: the tied constraints No-PATIENT-RelPron and MATCH are ranked above HIERARCHY. The ranking MATCH >> HIERARCHY fits with previous research of off-line studies that report case match to be more acceptable than case mismatch (Bausewein, 1991; Mewe, 2014, Vogel & Frisch, 2003, Vogel & Zugck, 2003, Vogel 2011).

Experiment 2

Experimental design and material

The experimental design was identical to the design of EXP 1. The cases of conditions and fillers were adjusted to the current case combination (nominative and accusative), see (13). The material was generated as outlined in EXP 1, but minor modifications concerning the adverb and the case of the NP following the relative pronoun were made. Specifically, the adverb was placed in front of the NP to clarify that it is not the NP's determiner that caused the effects of EXP 1 (position 2). In addition, nominative case is the default for subjects and double nominatives are ungrammatical in German (with few exceptions). Therefore, when the relative pronoun was nominative, the case marking of the following NP was changed to accusative (see (13) a. and d.). (13) illustrates the conditions with superscripts indicating sentence positions referred to during statistical analysis.

(13) Examples for the four conditions in EXP 2

a. nominative match (mat-nomnom)
Keine Geduld besaß ___, wer^{1} letztens^{2} den^{3} Sohn^{4} korrigiert^{5} hatte^{6}.
No patience had $\_\_\__{nom}$, who$_{nom}$ recently the$_{acc}$ son$_{acc}$ corrected has.
'No patience had, who the son has recently corrected.'

b. accusative match (mat-accacc)
Der Vater umarmte ___, wen^{1} letztens^{2} der^{3} Sohn^{4} korrigiert^{5} hatte^{6}.
The father hugged $\_\_\__{acc}$, whom$_{acc}$ recently the$_{nom}$ son$_{nom}$ corrected has.
'The father hugged, whom the son has recently corrected.'

c. hierarchy harmonic nominative-accusative mismatch (har-nomacc)
Keine Geduld besaß ___, wen^{1} letztens^{2} der^{3} Sohn^{4} korrigiert^{5} hatte^{6}.
No patience had $\_\_\__{nom}$, whom$_{acc}$ recently the$_{nom}$ son$_{nom}$ corrected has.
'No patience had, whom the son has recently corrected.'

d. hierarchy opposing accusative-nominative mismatch (opp-accnom)
Der Vater umarmte ___, wer^{1} letztens^{2} den^{3} Sohn^{4} korrigiert^{5} hatte^{6}.
The father hugged $\_\_\__{acc}$, who$_{nom}$ recently the$_{acc}$ son$_{acc}$ corrected has.
'The father hugged, who the son has recently corrected.'

Participants and procedure
96 monolingual German native speakers participated in this experiment under the same conditions as described in EXP 1. None of them took part in EXP 1. The procedure was identical to EXP 1 in order to keep the experiments comparable.

Results
For statistical analysis, the same statistical modelling approach as for EXP 1 was used; a linear mixed-effects model with the fixed factor CASE and random intercepts for participants and items (lmer(log(rt) ~ case + (1|participant) + (1|item)). Position 1 (relative pronoun) shows consistency with EXP 1 since it did not reveal a reaction time difference.

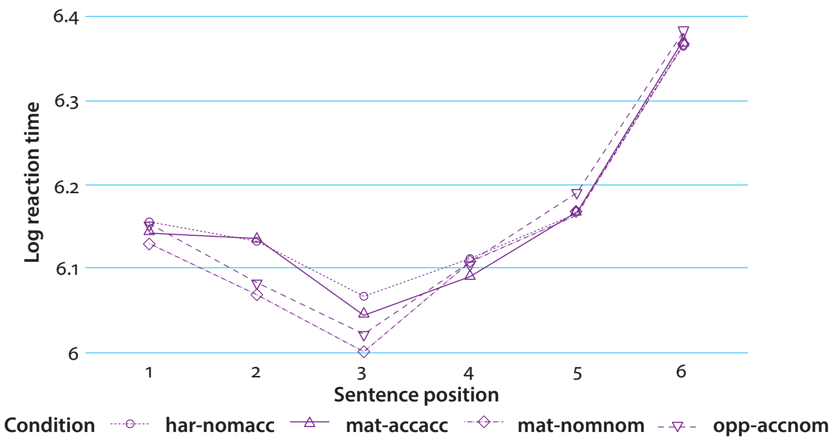

Figure 2. Logarithmised reaction times data on individually measured sentence positions

At position 2, the comparison between the case match conditions showed a significantly faster reaction time for the nominative match (mat-nomnom vs. mat-accacc; $t = -4.52$). The nominative match further revealed a reaction time advantage towards the hierarchy harmonic case mismatch (mat-nomnom vs. har-nomacc; $t = 4.14$). The accusative match was read significantly slower than the hierarchy opposing mismatch (mat-accacc vs. opp-accnom, $t = 3.64$). The case mismatches differed significantly from each other with the hierarchy opposing case mismatch having faster reaction times (har-nomacc vs. opp-nomacc; $t = 3.27$).

At position 3, the nominative case match was read significantly faster than the accusative case match (mat-nomnom vs. mat-accacc; $t = 3.03$) and the harmonic mismatch (mat-nomnom vs. har-nomacc; $t = 4.92$). The mismatch conditions showed a significant reaction time difference in favour of the opposing mismatch (har-nomacc vs. opp-accnom, $t = -3.54$).

Interim discussion for Experiment 2

Equivalently to EXP 1 the results will be interpreted using an OT-style analysis. Sentence position 3 is a post-case noun in EXP 1 but a case unambiguous article (*der/den*) in EXP 2 and EXP 3. The interpretation of the results for positions 2 and 3 will be provided in a single Table 3.

Table 3. OT-style analysis for positions 2 and 3

	No- PATIENT-RelPron	MATCH	HIERARCHY
→mat-nomnom			
mat-accacc	*		
har-nomacc	*	*	
?opp-accnom		*	*

Similarly to EXP 1 there is an intrinsic winner: nominative match (mat-nomnom). This condition does not violate any of the constraints. Thus, any ranking of the constraints would lead to the nominative match being the winner, rendering a ranking irrelevant. This is symbolised by missing vertical lines in Table 3. For position 2 most results can be captured by Table 3. Nominative match, as the intrinsic winner, was read significantly faster than the accusative match (mat-accacc) and the harmonic mismatch (har-nomacc). The comparison between the nominative match (mat-nomnom) and the opposing mismatch (opp-accnom) failed to reach statistical significance. This discrepancy between the results of the OT-style analysis and those of the statistical analysis is symbolised by a question mark in Table 3. However, there was a numerical reading times increase from nominative match to opposing mismatch. The opposing mismatch was read significantly faster than the remaining two conditions (mat-accacc and har-nomacc). This cannot be captured by an OT-style analysis, which yields mat-nomnom as the only winner.

Position 3 can be captured rather precisely in Table 3. As suggested by the table, the conditions accusative match (mat-accacc), harmonic mismatch (har-nomacc), and opposing mismatch (opp-accnom) did not reveal significant reading time differences to each other since they are intrinsic losers in the evaluation. Again, the comparison between nominative match (mat-nomnom) and opposing mismatch (opp-accnom) failed to reveal a statistically significant difference.

Recall that this is an OT-style analysis that should not be confused with traditional OT; thus, not all of the statistical results can be fully reflected in the Table 3 and the current approach provides a preliminary approximation.

Experiment 3

Experimental design and material

The experimental design was identical to the designs of EXP 1 and EXP 2. Similar to EXP 2 but different from EXP 1, sentence position 2 is an adverb and sentence position 3 is a case unambiguous article (*der/dem*). The cases of conditions and fillers were adjusted to the current case combination (nominative and dative), see (14). The material was generated as outlined in EXP 2:

(14) Examples for the four conditions in EXP 3

a. nominative match (mat-nomnom)
Keine Geduld besaß ___, wer[1] letztens[2] dem[3] Sohn[4] vertraut
No patience had $\_\_\__{nom}$, who$_{nom}$ recently the$_{dat}$ son$_{dat}$ trusted
hatte[6].
has.
'No patience had, who the son has promptly corrected.

b. dative match (mat-datdat)
Der Vater half ___, wem[1] letztens[2] der[3] Sohn[4] vertraut[5]
The father helped $\_\_\__{dat}$, whom$_{dat}$ recently the$_{nom}$ son$_{nom}$ trusted
hatte[6].
has.
'The father helped, whom the son has recently trusted.'

c. hierarchy harmonic nominative-accusaitve mismatch (har-nomacc)
Keine Geduld besaß ___, wem[1] letztens[2] der[3] Sohn[4] vertraut[5]
No patience had $\_\_\__{nom}$, whom$_{dat}$ recently the$_{nom}$ son$_{nom}$ trusted
hatte[6].
has.
'No patience had, whom the son has recently trusted.'

d. hierarchy opposing accusaitve-nominative mismatch (opp-accnom)
Der Vater half ___, wer[1] letztens[2] dem[3] Sohn[4] vertraut hatte[6].
The father helped $\_\_\__{dat}$, who$_{nom}$ recently the$_{dat}$ son$_{dat}$ trusted has.
'The father helped, whom the son has recently corrected.'

Participants and procedure

96 monolingual German native speakers participated in the experiment under the same conditions as described in EXP 1. None of them took part in EXP 1 or EXP 2. The procedure was similar to EXP 1 in order to keep the experiments comparable.

Results

For statistical analysis the same linear mixed-effects model as before was used with the fixed factor CASE and random intercepts by participants and items (lmer(log(rt) ~ case + (1|participant) + (1|item)). The sentence positions 1–6 are provided as indices in superscript on the German test material in (14 a–d).

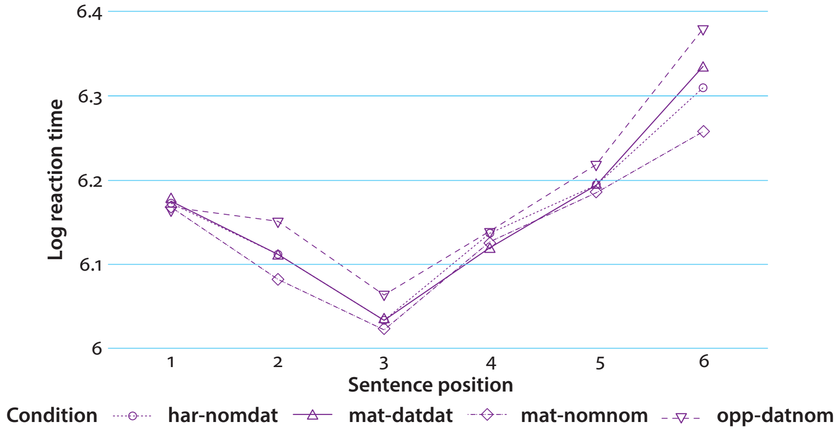

Figure 3. Logarithmised reaction times data on individually measured sentence positions

Position 2 (P2) and 3 (P3) revealed identical reaction time differences, and will be discussed together in the following. There was no significant reaction time difference between the match conditions (mat-nomnom vs. mat-datdat). The match conditions did not differ from the hierarchy harmonic mismatch (har-nomdat). The nominative match, the dative match and the hierarchy harmonic mismatch were read significantly faster than the hierarchy opposing mismatch. The comparison between the nominative match and opposing mismatch shows the largest difference (mat-nomnom vs. opp-datnom; P2: $t = -4.71$; P3: $t = -3.20$) followed by the comparison of the dative match and the opposing mismatch (mat-datdat vs. opp-datnom; P2: $t = 2.95$; P3: $t = 2.61$), followed by the comparison of the harmonic mismatch and opposing mismatch (har-nomdat vs. opp-datnom; P2: $t = 3.10$; P3: $t = 2.40$). Contrary to EXP 1 and EXP 2, position 6 (the spill-over region of the final verb) revealed reading time differences. The nominative match (mat-nomnom) was read significantly faster than the dative match (mat-nomnom vs. mat-datdat, $t = 3.25$). The nominative match was also read faster than the harmonic mismatch (mat-nomnom vs. har-nomdat, $t = 2.40$) and the opposing mismatch (mat-nomnom vs. opp-datnom, $t = 4.91$). The harmonic mismatch was read significantly faster than the opposing mismatch (har-nomdat vs. opp-datnom, $t = 2.52$).

Interim discussion for Experiment 3

Let us now interpret the results using the established OT-style analysis. The statistical results for sentence position 2 and 3 are identical and, thus, presented combined in table 4.

Table 4. OT-style analysis for position 2 and 3

	No-PATIENT-RelPron	MATCH	HIERARCHY
→ mat-nomnom			
→ mat-datdat			
→ har-nomdat		*	
opp-datnom		*	*

The statistical results are captured most fittingly by assuming two tied constraints, No-PATIENT-RelPron and MATCH, both ranked over HIERARCHY. The opposing mismatch (opp-datnom) was read significantly slower than the other conditions for sentence position 2 and 3. Thus, the significant reading time differences between the match conditions and the opposing mismatch condition is accurately captured by Table 4. Although the harmonic mismatch (har-nomdat) violates MATCH, it does not result in a statistically significant difference; it does show a numerical reading time increase compared to the match conditions. The results of position 2 and 3 of the current experiment reveal one crucial finding: Since the tie between NO-PATIENT-RelPron and MATCH does not lead to the elimination of any candidate, it is HIERARCHY that is responsible for the elimination of the hierarchy opposing mismatch. This is in line with the statistical results: opp-datnom is the only condition that is read significantly slower than the other conditions. Importantly, whereas the interpretation of the results of position 2 and 3 in EXP 1 and EXP 2 relied on all constraints, the results of position 2 and 3 in EXP 3 can be explained solely by this constraint. This is an important finding because it ties in with previous results from off-line data centring around the hierarchy rule this constraint is derived from.

EXP 3 is remarkable in yet another aspect: a reading times difference at position 6 was found. At position 5 the final and disambiguating verb is encountered which led to reading times differences at position 6, the verb's spill-over region. At this point the parser must integrate the final verb. For the conditions with an initial dative (mat-datdat and har-nomdat), the assumption of a dative-experiencer or –possessor cannot be kept since there are no verbs in the test material licencing such thematic roles in the conditions. Hence, at position 5 the need arises to reanalyse the dative from an experiencer/possessor (an agentive role) to a patient role. This is convincingly reflected in the results in the spill over region of position 5, interpreted using an OT-style analysis given in Table 5.

Table 5. OT-style analysis for position 6

	No-PATIENT-RelPron	MATCH	HIERARCHY
→mat-nomnom			
mat-datdat	*		
har-nomdat	*	*	
opp-datnom		*	*

Similarly to EXP 1 (position 3, see Table 2) and 3 (position 2 and 3, see Table 4) tied constraints are assumed; that is, No-PATIENT-RelPron and MATCH. Again, for the nominative match (mat-nomnom) a ranking is irrelevant because it does not violate any constraints. Since at position 6 the final verb was integrated into the sentence analysis of the parser No-PATIENT-RelPron gains relevance. The locally tied constraints lead to two evaluations. In the first evaluation, that is No-PATIENT-RelPron >> MATCH, the match nominative match (mat-nomnom) and the opposing mismatch (opp-datnom) would be the winners. In the second evaluation, that is MATCH >> No-PATIENT-RelPron, the two match conditions (mat-nomnom and mat-datdat) would be the winners. The nominative match (mat-nomnom) is the intrinsic winner of the evaluation. This accurately captures the results where the nominative match was read faster than the other conditions. The dative match (mat-datdat) and the harmonic mismatch (har-nomdat) failed to reveal a significant reading time difference; since in both cases the thematic role of the initial dative is a patient, it violates No-PATIENT-RelPron. The HARMONIC mismatch (har-nomdat) and the opposing mismatch (opp-datnom) violate MATCH. Due to the locally tied constraints No-PATIENT-RelPron and MATCH the violation of No-PATIENT-RelPron of the harmonic mismatch (har-nomdat) might not be decisive. However, the violating of HIERARCHY of the opposing mismatch (opp-datnom) explains the longer reading times compared to the harmonic mismatch (har-nomdat). For the time being, it is unclear why this does not hold for the comparison between the dative match (mat-datdat) and the opposing mismatch (opp-datnom).

3. General discussion

Let us recapitulate the results of our study. In EXP 1, which tested the case constellation accusative-dative, the dative match (mat-datdat) was read significantly faster than all other conditions on position 2 (article), the first spillover region after the relative pronoun (mat-datdat << mat-accacc, har-accdat, opp-datacc; << abbreviates shorter reading times in the following). On position 3 (noun), the dative match was also read significantly faster than the slowest condition: the hierarchy opposing mismatch (opp-datacc). There were no other significant reading times differences (mat-datdat << opp-datacc; mat-accacc = har-accdat = opp-datacc).

In EXP 2, which focussed on the case combination nominative-accusative, on position 2 (adverb) the nominative match (mat-nomnom) and the hierarchy opposing mismatch (opp-accnom) were read significantly faster than the accusative match (mat-accacc) and the hierarchy harmonic mismatch (har-nomacc). The first two and latter two conditions did not reveal significant reading time differences relative to each other (mat-nomnom, opp-accnom << mat-accacc, har-nomacc). On position 3 (article) faster reading times of the nominative match (mat-nomnom) against the accusative match (mat-accacc) and the hierarchy harmonic mismatch (har-nomacc) were found.

EXP 3, which was devoted to the case constellation nominative-dative, featured identical reading time differences for position 2 (adverb) and position 3 (article): the hierarchy opposing mismatch (opp-datnom) was read significantly slower than the other conditions. The other conditions did not reveal significant reading time differences (mat-nomnom, mat-datdat, har-nomdat << opp-datnom). Interestingly, position 6 also revealed significant reading time differences. It is the spill-over region of the final verb that disambiguates the initial dative relative pronoun from an assumed experiencer- or possessor-dative to a patient-like dative. For initial patient-like datives the default word order does not hold. Thus, at this point the violation No-PATIENT-RelPron becomes apparent. The corrected interpretation of the initial dative elicits longer reading times for conditions with intital dative relative pronouns (mat-datdat and har-nomdat) compared to the nominative match (mat-nomnom). However, the hierarchy opposing mismatch (opp-datnom) remains the slowest condition due to its violation of HIERARCHY (mat-nomnom < mat-datdat, har-nomdat < opp-datnom).

The OT-style modeling of the statistical results has proved to be quite fruitful. It has shown that all three constraints are active in the incremental self-paced reading of free relative clauses: relative pronouns with a patient role in the accusative are avoided (No-PATIENT-RelPron), the case of the covert head matches the case of the relative pronoun (MATCH) and the case of the covert head is higher on the case hierarchy than the case of the relative pronoun (HIERARCHY).

Let us start the general discussion with the HIERARCHY constraint. Its relevance is particularly evident in EXP 3 on position 2 and 3, where all stimuli obey No-PATIENT-RelPron. Due to the fact that this constraint seems to be tied to MATCH, it is only HIERARCHY that bans the hierarchy opposing mismatch (opp-datnom), the only type of stimulus that has elicited significantly longer reading times against the other stimuli at position 2 and 3. On position 6, HIERARCHY distinguishes the mismatch conditions (har-nomdat and opp-datnom), capturing the reading time difference between these conditions accurately.

In off-line experiments (Bausewein, 1991; Mewe, 2014; Vogel & Frisch 2003; Vogel & Zugck, 2003; Vogel, 2011) case matches, i.e. the mat-conditions, were judged to be more acceptable than case mismatches that were in harmony with the HIERARCHY constraint, i.e. the har-conditions. This means that MATCH is a stronger constraint than HIERARCHY. The results of the present study are in line with this assumption. Promoting HIERARCHY to a higher rank would not improve the explanation of the overall online data.

Finally, let us turn to the constraint against relative pronouns with a patient role in an objective case: No-PATIENT-RelPron. Its relevance is manifested in EXP 1 and EXP 2, where it is necessary to explain the statistical data appropriately, the shortest reading times for the dative case match (mat-datdat) against the accusative match (mat-accacc), in particular. Neither MATCH nor HIERARCHY captures this result. Further, results on position 6 in EXP 3 strongly support this constraint since the wrongly interpreted experiencer- or possessor-dative becomes apparent and reanalysis towards a patient-like dative elicits longer reading times. This is remarkable in two respects. First, it shows that the subject-first strategy, a strategy mainly targeted towards the grammatical object and inspired from previous research (Keenan & Comrie, 1977; Sheldon, 1974; MacWhinney & Pléh, 1988) is inappropriate for German on-line data, as it would not distinguish dative match from accusative match unequivocally, as explained in the introduction. Therefore, in our tables we have replaced the grammatical object function by the semantic patient role (No-PATIENT-RelPron). Second, No-PATIENT-RelPron has not been shown to guide the off-line processing of free relative clauses (Bausewein, 1991; Mewe, 2014; Vogel & Frisch, 2003; Vogel & Zugck, 2003; Vogel 2011). In other words, there was no manifest advantage for a dative match in off-line experiments.

Let us try to explain this asymmetry. In a self-paced reading experiment, participants process cases and thematic roles incrementally before the sentence final verb is encountered – similarly to what has been reported for incremental sentence comprehension with ERP measurements (cf. Bornkessel, 2002; Bornkessel et al., 2003). When encountering a dative relative pronoun participants have no reason to assume that this dative encodes a patient. As described in the introduction, datives quite frequently encode an agentive role in German. In incremental interpretation the dative match is therefore clearly in advantage against an accusative match. It is only when the final verb is encountered that the patient interpretation of the dative becomes obvious since all verbs in our study – and all verbs in the off-line experiments – select a patient-like dative: e.g. *helfen* (help), *applaudieren* (applaude), *vertrauen* (trust), *assistieren* (assist), *dienen* (serve). The results of EXP 3 support this by showing a reanalysis effect on position 6 for conditions with initial dative relative pronouns, where the incorrect early interpretation of an experiencer-dative or a possessor-dative is corrected to a patient-like dative. By contrast, in an off-line experiment, the whole sentence including the verb is evaluated globally; a local advantage for a relative pronoun with a certain thematic role is not found.

However, the three constraints under study cannot capture all the data. First, in EXP 2 the hierarchy opposing mismatch with an accusative covert head and a nominative relative pronoun (opp-accnom) failed to reach a significant reading time difference compared to the nominative match (mat-nomnom). This cannot be captured by an OT-style analysis, which yields mat-nomnom as the only winner.

Second, another finding that the constraints cannot fully capture occurred in EXP 1. The hierarchy opposing mismatch (opp-datacc) was the only condition to reveal a significant reading time difference against the dative match (mat-datdat) on position 3, although according to the OT-style analysis the dative match is the intrinsic winner of the evaluation. However, the numeric gradation between the conditions is reflected in the OT-style analysis and fully captured on position 2.

Finally, another open issue is the exact nature of the constraint interaction. In some cases the ranking of the constraints seems to be irrelevant, as in EXP 1 at position 2 and EXP 2 at positions 2 and 3. In these instances there is only one "winner" and it would win the competition under the assumption of any strict ranking. Yet in other cases, e.g. EXP 1 position 3 and partially, i.e. wrt to MATCH and No-PATIENT-RelPron, in EXP 3 at positions 2 and 3, two or three constraints are tied in an equal rank. This is the method of capturing multiple "winners" in OT, i.e. multiple optimal candidates results from tied constraints.

In sum, our series of experiment has provided converging evidence that the comprehension of free relative clauses is guided by three principles (HIERARCHY, MATCH and No-PATIENT-RelPron) whose ranking can be described adequately with an OT-style approach. The present data also highlight that, in order to fully capture the processing characteristics of free relative clauses, multiple experimental methods should be used. In particular, self-paced reading as a behavioural measure of incremental sentence processing has provided evidence that the principles above differentially affect the processing of clausal constituents within the free relative clause, which may in fact be masked in off-line acceptability ratings.

4. Conclusions

Previous research has used off-line methods such as acceptability judgements or corpus counts to examine the processing of free relative clauses in German, and has provided converging evidence for the applicability of the hierarchy rule strategy. The present study investigated to what extent these findings are corroborated by findings from on-line sentence processing as participants read free relative clauses in the self-paced reading paradigm. We tested case combinations of nominative vs. accusative, nominative vs. dative and accusative vs. dative case across three experiments.

Overall, the results suggest that, for self-paced reading, particularities of incremental sentence processing have to be taken into account to fully capture the data. Previous off-line studies have mostly focussed on the hierarchy rule as the crucial factor to explain the findings. We considered two important commonalities between free relative clauses and headed relative clauses, i.e. the parallel syntactic function strategy and the subject-first strategy. Especially the latter has hitherto been neglected in research on free relative clauses in German. For attributive relative clauses, the parallel syntactic function strategy predicts less processing cost for case matches in general, while the subject-first strategy, reinterpreted from targeting grammatical function to targeting thematic roles, predicts processing ease for relative pronouns compatible with a subject or agent interpretation. Both predictions were supported by our empirical findings for free relative clauses. In order to model the three strategies (hierarchy rule strategy, parallel syntactic function strategy, subject-first strategy) systematically, we translated them into constraints (HIERARCHY, MATCH, No-PATIENT-RelPron) for an OT-style analysis assuming they interact with each other. This approach can capture results from previous off-line studies, but, importantly, also accounts for accusative-nominative mismatches that have shown patterns deviant from predictions of the hierarchy rule strategy alone. We argue that it is the constraint No-PATIENT-RelPron in particular that can explain the inconclusive results regarding this hierarchy opposing case mismatch. Thus, the OT-style analysis is superior in explaining the overall pattern of findings across off-line methods and self-paced reading, yet may still benefit from further research using on-line measures with an even better temporal resolution than provided by self-paced reading.

References

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/S0305000909990080

Bausewein, K. (1991). Haben kopflose Relativsätze tatsächlich keine Köpfe? In G. Fanselow & S. Felix (Eds.), *Strukturen und Merkmale grammatischer Kategorien* (pp. 144–158). Tübingen: Narr.

Bornkessel-Schlesewsky, I. (2002). *The Argument Dependency Model: A neurocognitive approach to incremental interpretation* (Unpublished doctoral dissertation). Max Planck Institute of Cognitive Neuroscience, Leipzig.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). Minimality as vacuous distinctness: Evidence from cross-linguistic sentence comprehension. *Lingua*, 119(10), 1541–1559. https://doi.org/10.1016/j.lingua.2008.03.005

Bornkessel-Schlesewsky, I., Schlesewsky, M., & Friederici, A. D. (2003). Eliciting thematic reanalysis effects: The role of syntax-independent information during parsing. *Language and Cognitive Processes*, 18(3), 269–298. https://doi.org/10.1080/01690960244000018

Crocker, M. W. (1994). On the nature of the principle-based sentence processor. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 245–266). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dröge, A., Maffongelli, L., & Bornkessel-Schlesewsky, I. (2014). Luigi piace a Laura? Electrophysiological evidence for thematic reanalysis with Italian dative object experiencer verbs. In A. Bachrach, I. Roy, & L. Stockall (Eds.), *Structuring the argument: Multidisciplinary research on verb argument structure* (pp. 83–118). Amsterdam: John Benjamins.

Fanselow, G., Schlesewsky, M., Cavar, D., & Kliegl, R. (1999). *Optimal parsing: Syntactic parsing preferences and optimality theory*. In *Rutgers Optimality Archive (ROA) 367*. Rutgers State University of New Jersey.

Groos, A., & van Riemsdijk, H. (1981). Matching effects in free relatives: A parameter of core grammar. In A. Belletti (Ed.), *Theory of markedness in generative grammar. Proceedings of the 1979 GLOW Conference* (pp. 171–216). Pisa: Scuola Normale Superiore.

Keenan, E., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8, 63–99. https://doi.org/10.2307/413321

Kirby, S. (1998). *Function, selection and innateness: The emergence of language universals*. Oxford: Oxford University Press.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mewe, J. (2014). *Kasusmatchting bei freien Relativsätzen im Deutschen – Eine experimentelle Untersuchung* (Unpublished MA thesis). University of Cologne.

Newmeyer, F. J. (2005). *Possible and probably languages. A generative perspective on linguistic typology*. Oxford: Oxford University Press. https://doi.org/10.1017/S0022226706254586

Pittner, K. (1991). Freie Relativsätze und die Kasushierarchie. In E. Feldbusch, R. Pogarell, & C. Weiß (Eds.), *Neue Fragen der Linguistik. Akten des 25. Linguistischen Kolloquiums, Paderborn 1990. Band 1*. (Linguistische Arbeiten 270) (pp. 341–347). Tübingen: Niemeyer.

Pittner, K. (2003). Kasuskonflikte bei freien Relativsätzen - Eine Korpusstudie. *Deutsche Sprache*, 31(3), 193–208.

Primus, B. (1999). *Cases and thematic roles - Ergative, accusative and active*. Tübingen: Niemeyer. https://doi.org/10.1515/9783110912463

Primus, B. (2015). Grammatical relations. In T. Kiss & A. Alexiadou (Eds.), *Syntax – Theory and analysis. An international handbook* (2nd ed., pp. 219–247). Berlin: De Gruyter.

Prince, A., & Smolensky, P. (2002). *Optimality Theory. Constraint interaction in generative grammar*. Hoboken, NJ: Wiley & Sons.

Schlesewsky, M. (1997). *Kasusphänomene in der Sprachverarbeitung. Eine Studie zur Verarbeitung von kasusmarkierten und Relativsatzkonstruktionen im Deutschen* (Unpublished doctoral dissertation). University of Potsdam.

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, 13, 272–281. https://doi.org/10.1016/S0022-5371(74)80064-2

van Riemsdijk, H. (2006). Free relatives. In M. Everaert & H. C. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax* (Vol. II, pp. 336–378). Oxford: Blackwell. https://doi.org/10.1002/9780470996591.ch27

Vogel, R. (2011). Disagreement, variation, markedness and other apparent exceptions. In H. J. Simon & H. Wiese (Eds.), *Expecting the unexpected: Exceptions in grammar* (pp. 339–359). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110219098.339

Vogel, R., & Frisch, S. (2003). The resolution of case conflicts. A pilot study. In S. Fischer, R. van de Vijver, & R. Vogel (Eds.), *Linguistics in Potsdam, Vol. 21, Experimental studies in linguistics 1* (pp. 91–103). Potsdam: Potsdam University Press. https://doi.org/10.1515/9783110219098.339

Vogel, R., Frisch, S., & Zugck, M. (2006). Case matching. An empirical study on the distinction between abstract case and case morphology. *Linguistische Berichte*, 208, 357–383.

Zugck, M., & Vogel, R. (2004). Counting markedness. A corpus investigation on German free relative constructions. In S. Fischer, R. van de Vijver, & R. Vogel (Eds.), *Linguistics in Potsdam, Vol. 21: Experimental studies in linguistics 1* (pp. 105–122). Potsdam: University Press.

Integrating the filler

Evidence from double object constructions in Greek relative clauses

Anastasia Paspali
Humboldt University of Berlin

This study investigates filler integration by adult native speakers of Greek. Object Relative Clauses (RCs) including ditransitives were tested in order to address (a) whether Greek adults show reactivation of the filler at the gap as proposed by the Trace Reactivation Hypothesis (TRH), and (b) the psychological reality of two theoretical analyses of ditransitives in Greek. The first analysis proposes that the base-generated order is direct object (DO)–indirect object (IO), where the DO c-commands the IO, and the second analysis proposes that the basic word order is IO–DO, where the IO c-commands the DO. Cross-Modal Priming (CMP) (Swinney et al., 1979) and Probe-Classification During Reading aloud (Dekydtspotter et al., 2010) were employed. Trace reactivation effects confirm the TRH only in the CMP task and only when adopting the IO–DO order as basic.

Keywords: filler integration in Greek, cross-modal priming, probe classification during reading aloud, recipient prepositional constructions

1. Introduction

A considerable body of online comprehension studies focuses on the processing of traces and filler integration cross-linguistically. Since the early studies (see e.g., Nicol, 1993; Traxler & Pickering, 1996) and more recent ones (see e.g., Nakano, Felser, & Clahsen, 2002; Roberts, Marinis, Felser, & Clahsen, 2007), a central question addressed the psychological reality of traces and the way in which the parser mentally reactivates them.

Two processing accounts have been proposed on filler integration: the Trace Reactivation Hypothesis (TRH) and the Direct Association (DAH) Hypothesis (Pickering & Barry, 1991). According to the TRH, filler integration is modulated

https://doi.org/10.1075/lald.64.02pas

by syntactic traces in line with the generative-transformational framework. For example, while processing a sentence as in (1), where the relative pronoun *to which* is the filler, and t_i represents its gap[1] (the position where the filler originates before movement), the parser keeps the filler in working memory. On encountering the gap, the parser recalls and regenerates (e.g., integrates) the semantic and grammatical features of the filler in order to interpret the sentence.

(1) John saw the camel to which$_i$ the monkey gave the nice present t_i at the party.

On the other hand, the DAH disputes the psychological reality of traces (Pickering & Barry, 1991, pp. 229–230), deriving its background from trace-free grammars. This processing account supports a direct linking between the filler and its subcategorizer (i.e., the verb) which is driven by semantic rather than syntactic information. Namely, it predicts that the filler is integrated into the structure when the parser encounters the verb which assigns a thematic role to the filler; hence, reactivation effects are expected immediately after the verb (i.e., *gave* in (1)).

As noted in the literature, it is highly difficult to empirically dissociate these accounts in head-initial languages (Clahsen & Featherston, 1999, p. 410; Felser & Roberts, 2007, p. 11; Marinis, Roberts, Felser, & Clahsen, 2005, p. 55) because the trace of the DO (2) immediately follows the verb and both accounts predict that effects occur at the same point.

(2) John saw the ring$_\mathrm{i}$ which Mary stole t_i at the party.

Indeed, studies have provided evidence for both of these accounts (e.g., Nicol et al., 1994; Traxler & Pickering, 1996). For this reason, Felser and Roberts (2007, p. 11) point out that more reliable evidence on trace reactivation effects at the point of the gap can be drawn from studies with head-final languages, indirect objects dependencies, subject RCs, and dependencies spanning more than one clause. Overall, most studies that empirically dissociated the verb from the gap, confirmed the TRH cross-linguistically both in head-final and head-initial languages and irrespective of the experimental task employed.

The aims of the current study are the following:

a. to explore the psychological reality of traces by testing the TRH;[2] following the previous literature, the subcategorizing verb was dissociated from the

1. The terms "gap" and "trace" are used interchangeably in the present study and their representation is the following: t_i.

2. As an anonymous reviewer pointed out, only the position of the gap and not the position of the subcategorizing verb was tested in the current study, following Felser & Roberts (2007) and

gap (see Felser & Roberts, 2007 and references therein) by using ditransitive constructions,

b. to compare two methodological paradigms that have been used in the literature to test filler integration: Cross-modal Priming and Probe Classification During Reading, and
c. to empirically test two competing syntactic analyses about the base-generated and the derived order of arguments in prepositional double object constructions in Greek; given that the position of the trace depends on which argument of a ditransitive construction is considered hierarchically higher, we manipulated the stimuli according to these competing syntactic analyses.

In what follows, the methodological background is sketched out and the two syntactic analyses of double object constructions in Greek are described. Then, the experiments of the current study are reported followed by discussion.

2. Methodological background

2.1 Cross-modal priming

In this study, two distinct experimental paradigms are applied: Cross-modal Priming (CMP) (Swinney et al., 1979) and Probe Classification During Reading aloud (PCDR) (Dekydtspotter et al., 2010). Both tasks have been used to explore the psychological reality of traces.

The procedure of a CMP is the following: participants listen to sentences in natural speed through headphones while they are seated in front of a computer screen. While listening to the sentences, a stimulus (word/picture) appears on the display. As soon as they see the stimulus, participants have to make a lexical decision or a picture classification related to this stimulus (e.g., to categorize the picture as animate or inanimate) by pressing one of two keys on the keyboard or a response box as fast as possible. Reaction Times (RTs) are measured between the time the stimulus appears on the computer screen and the time participants

Miller (2015) among others. The primary goal of the current study is to explore whether the parser mentally reactivates traces in Greek, in other words, whether traces play a role in sentence processing and not whether the predictions of the DAH at the point of the subcategorizing verb are confirmed. Thus, testing the positions of the subcategorizing verb, where the DAH predicts effects and the TRH does not, may confirm the DAH but cannot contribute to our primary aim; the psychological reality of traces. On the other hand, testing the position of the gap, where TRH predicts trace reactivation effects but the DAH does not, will allow us to draw conclusions about the processing of traces (see Clahsen & Featherston, 1999 p. 420 for further discussion on the topic).

press the key. Comprehension questions are included after each trial to verify the participants' attention (see e.g., Marinis, 2010; 2018). The CMP must be carefully designed and implemented (Nicol et al., 1994, p. 2) and certain factors should be controlled for, such as the length of antecedents as well as their frequency.[3]

The rationale of this task is that: (a) priming effects (shorter RTs) are expected at the point of the hypothesized trace when the picture matches the antecedent of the filler, compared to an unrelated picture at the same position, due to activation of the semantic and grammatical features of the antecedent at the point of the gap; and (b) reactivation effects (shorter RTs) should be position-specific, that is, shorter RTs are expected for the picture that matches the antecedent of the filler at the point of the gap compared to a control position. Thus, the most important indication of reactivation effects at the point of the gap, is a significant interaction between the two factors, namely the position that the picture appears (gap position or pre-gap position) and the type of the picture (identical to the filler or unrelated).

The CMP has been criticized in the literature (McKoon et al., 1994; Miller 2015) for the extent to which it is accurate in measuring trace reactivation effects (Miller, 2015; Dekydtspotter et al., 2010). Miller (2015) compared the CMP with the PCDR (see Probe Classification During Reading in this article for the procedure of the latter task) by using the same experimental materials in the same groups of participants. The results showed trace reactivation effects with the PCDR but not with the CMP.

2.2 Probe classification during reading

The second method employed in this study is the PCDR (Dekydtspotter et al., 2010). In this task, participants read sentences aloud segment by segment in a forced-paced way. While reading, a picture appears on the center of the display and they have to categorize it as animate or inanimate as fast as possible by pressing one of two keys on the keyboard while RTs are measured. The rationale is the same as in CMP.

Dekydtspotter et al. (2010) and Miller (2015) consider this task less demanding than the CMP. In their task, no comprehension questions were included and the stimuli were presented in a single modality (only visually). However, results of previous studies using this method are not conclusive; indeed, the results in Dekydtspotter & Miller (2013, second experiment), Dekydtspotter & Miller (2009), Dekydtspotter et al., (2010), and Miller (2013), do not present clear evidence of trace reactivation effects in NSs because the significant interaction

3. See Marinis (2018) for a concise overview of the cross-modal priming methodology.

between position and type of picture was absent. On the other hand, Dekydtspotter & Miller (2013, first experiment), Miller (2015), and Miller (2011) found trace reactivation effects in the same languages (namely, English and French).

Thus, the literature has shown mixed results so far with respect to the sensitivity of each task on capturing trace reactivation effects, especially for PCDR. CMP has been used extensively and most studies have revealed trace reactivation This reflects that CMP is a sensitive measure of real-time sentence comprehension when it is carefully designed and implemented (Marinis, 2018), as opposed to what Miller (2015) found. On the other hand, PCDR has been used less frequently and has revealed a less clear picture of results. Thus, in the current study, we further explore these methods by using the same design and materials, both for reasons of consistency and comparability.

3. Argument structure in Greek ditransitives

Greek includes a variety of ditransitive constructions (Holton, Mackridge, & Philippaki, 1997) where the IO can surface either as Prepositional Phrase (PP) or as Determiner Phrase (DP). There are two alternative structures for IOs which denote recipients in Greek; prepositional IOs marked with the preposition *se* and with morphological accusative case and IOs in genitive case. In our experiments only prepositional IOs were used. In this structure (3), the DO denotes the theme and the IO denotes the recipient. The construction can surface either as DO–IO (3a) or as IO–DO (3b).

(3) a. o Jánis éðose ta lulúðja s- ti Maria.
The$_{\text{NOM}}$ John$_{\text{NOM}}$ gave the$_{\text{ACC}}$ flowers$_{\text{ACC}}$ to-the$_{\text{ACC}}$ Mary$_{\text{ACC}}$
'John gave the flowers to Mary.'
b. o Jánis éðose s- ti Maria ta lulúðja.
The$_{\text{NOM}}$ John$_{\text{NOM}}$ gave to-the$_{\text{ACC}}$ Mary$_{\text{ACC}}$ the$_{\text{ACC}}$ flowers$_{\text{ACC}}$
'John gave Mary the flowers.'

There are three alternative analyses according to Anagnostopoulou (2005) that may account for the order of arguments in *se*-PP recipient constructions. The first analysis suggests that the DO is at a higher position dominating the IO; hence the DO–IO order is considered the base-generated order in Greek and the IO–DO is derived by A-scrambling of the PP over the DP. The second analysis proposes the opposite scenario, where IO–DO is the basic order and the IO dominates the DO. The third analysis draws from binding facts and suggests that both orders can be base-generated.

3.1 DO–IO as basic order

The DO–IO as base-generated order is supported by Michelioudakis (2011), who applies certain tests proposed by Barss & Lasnik (1986) to further explore c-command relations to all possible surface orders in Greek. He concludes that the prepositional IOs should always be kept distinct from IOs with DPs.[4] He excludes the possibility that the IO–DO can be basic in *se*-PP recipient constructions mainly for Case-related reasons; if the IO asymmetrically c-commands the DO, it must necessarily bear an active Case feature. This claim is problematic since PPs are not considered to bear active Case (p. 97).

3.2 IO–DO as basic

Alternatively, the IO–DO order has been argued to be the base-generated order in Greek as well (Georgala, 2012; Georgala & Whitman, 2009; and Bowers & Georgala; 2007; Goergiafentis & Lascaratou (2007). Georgala (2012), Georgala and Whitman (2009), and Bowers and Georgala (2007) point out that if both orders are considered base-generated, certain theoretical aspects are undermined, such as the Uniformity of θ-Assignment Hypothesis. Anagnostopoulou (2005) proposes that both orders might be basic based on Marantz (1993), who argues that there are certain thematic roles such that it does not matter where the one is merged relative to the other. More specifically, Instrument, Affected object Locative and Inalienable Possessor are affected simultaneously in the same event as the Theme and may be lower or higher than the Theme. However, as Georgala (2012) points out, potential Recipient/Goal[5] is not one of the thematic roles Marantz (1993) mentions.

Georgala (2012) further explores the hierarchy of arguments in prepositional IO-recipient structures by adopting the syntactic architecture of applicative constructions along the lines of Marantz (1993) and she shows that the IO–DO structure is basic by employing certain syntactic tests such as depictive stranding as well as idioms. Within the framework of applicative constructions, the light verb head, which is called Applicative (Appl), is charged with two conceptually different tasks, as Anagnostopoulou (2013) points out: syntactically licensing an extra object and assigning a thematic role. Georgala (2012) argues that the applicative projection is uniformly above the lexical VP in line with Pylkkänen (2002). The

4. He also extends his view that either the DO-IO order is basic or that both orders are base-generated in line with Anagnostopoulou (2005).

5. The literature of goal ditransitive constructions makes a distinction between the thematic role of Recipient goal and the thematic role of Locative goal (see Marantz, 1993 and Georgala, 2012 for further discussion).

EPP (Extended Projection Principle) on Appl, triggers movement of the IO to the specifier position of the Applicative Phrase (Spec, ApplP) and values its Case feature. Georgala (2012) first presents evidence that recipient constructions with genitive case and *se*-PP recipient constructions in Greek are generated in the same syntactic positions, namely [Spec, VP]. However, they exhibit different types of movement: Short Object Shift (A-scrambling) vs. A-bar movement.

Georgala mainly presents evidence from depictive stranding structures as in (4).

(4) a. éðosa [tin bira]$_i$ s- ti Léna zesti$_i$.
Give$_{\text{PAST-1SG}}$ the$_{\text{ACC}}$ beer$_{\text{ACC}}$ to-the$_{\text{ACC}}$ Lena$_{\text{ACC}}$ warm$_{\text{ACC}}$
'I gave the beer to Lena warm.'

b. *éðosa [tin bira]$_i$ tis Lénas zesti$_i$.
Give$_{\text{PAST-1SG}}$ the$_{\text{ACC}}$ beer$_{\text{ACC}}$ the$_{\text{GEN}}$ Lena$_{\text{GEN}}$ warm$_{\text{ACC}}$
'I gave the beer to Lena warm.'

Depictive stranding in (4a) indicates that the DO–IO object is derived by movement; According to Georgala, this is a Short Object Shift, an A-scrambling of the accusative DP to [Spec, vP].

Goergiafentis and Lascaratou (2007) also propose that the IO (whether realized as a PP or a DP) occupies a structurally higher position than that of the DO, and thus the DO is c-commanded by the IO, in line with Georgala (2012). They base their analysis on lexico-syntactic criteria, which reveal that the asymmetry between the IO and the DO, which has been observed in many languages, is also valid in Greek. They use evidence from scope and binding, theta marking and idioms, Object-Verb 'composites', deletion and case marking and passivisation.

To sum up, different syntactic analyses have been proposed for the hierarchical order of arguments in Greek ditransitive verbs. In this study, we focus on prepositional recipient ditransitives and we explore the issue experimentally.

4. The present study

The present study addresses (a) filler integration in Greek by testing the TRH, (b) the methodological aspects of testing filler integration by employing two distinct experimental paradigms, and (c) the psycholinguistic reality of the two competing syntactic analyses of prepositional double object constructions in Greek. Experiment 1 and Experiment 2 explore filler integration using IO RCs, while at the same time the DO-IO order is adopted as the base-generated order in Greek ditransitives. Consequently, the trace of the IO, which is the filler, is posited at a hierarchically lower position compared to the DO (Figure 1).

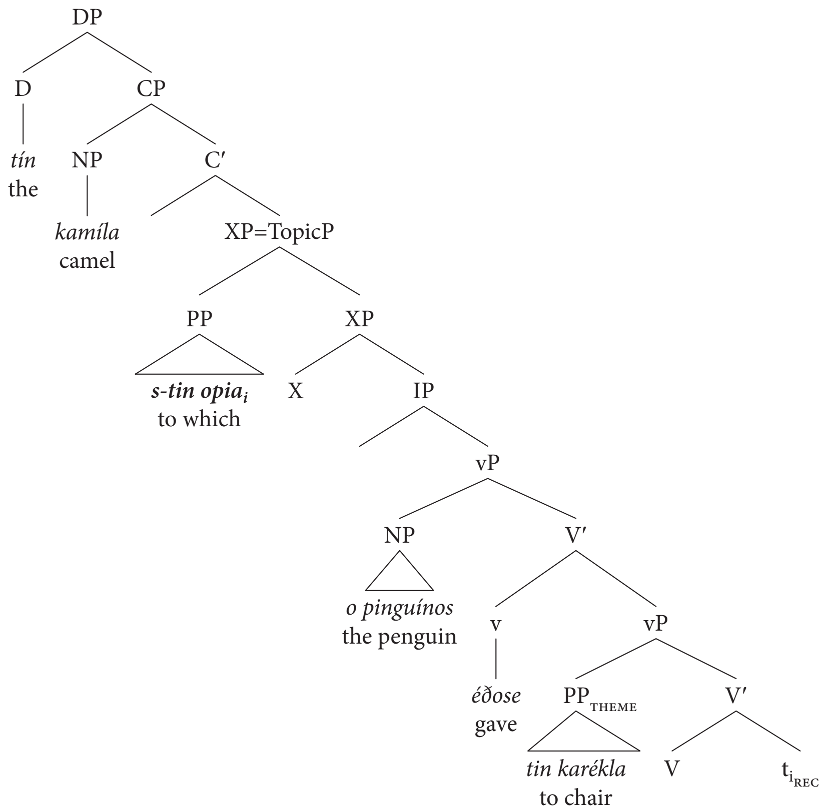

Figure 1. The base generated position of the DO dominates the IO in an IO RC (Experiments 1 & 2)[6]

While the same materials are used in both experiments, two different experimental methods are employed; Experiment 1 employs a PCDR task and Experiment 2 a CMP task. Experiments 3 and 4 adopt the view that the IO–DO is the base-generated order in Greek. Hence, the trace of the DO, which is the filler, is posited at a lower position compared to the IO object (Figure 2).

6. Here, I follow Alexiadou (1997: 20) on the derivation of Greek relative clauses (RCs) introduced with the pronoun *o opoios, -a, -o*. Alexiadou adopts a raising analysis for RCs, following Kayne (1994). According to Alexiadou the movement of the NP to Spec, CP is obligatory, so that the external Determiner can receive its NP complement. More specifically, the Head noun of the RC is generated inside the RC and is subsequently raised to the Spec, CP. Specifically, for RCs introduced with the pronoun *o opoios, -a, -o,* the NP movement to Spec,CP contains two steps. Step I involves fronting of the (prepositional) relative pronoun to the Spec of a lower projection of CP, which Alexiadou calls Topic Phrase. Step II takes the Head noun out of the relative expression, and moves it into the Spec,CP, a position which only allows a Head.

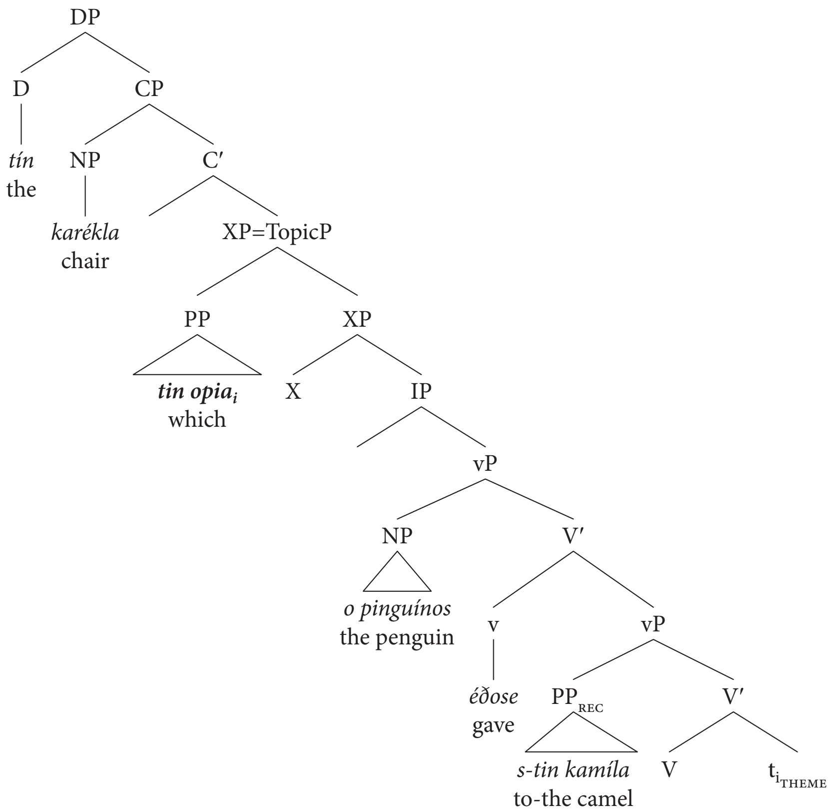

Figure 2. The base-generated position of the DO is dominated by the IO in a DO RC (Experiments 3 & 4)

The use of both IO RCs (Experiments 1 and 2) and DO RCs (Experiments 3 and 4) allows us to test the theoretical analyses on ditransitives mentioned above while keeping at the same time the position of the gap empirically dissociated from the verb. This is the reason we did not exploit IO RCs in Experiments 3 and 4 (Figure 3), given that the trace immediately follows the verb.

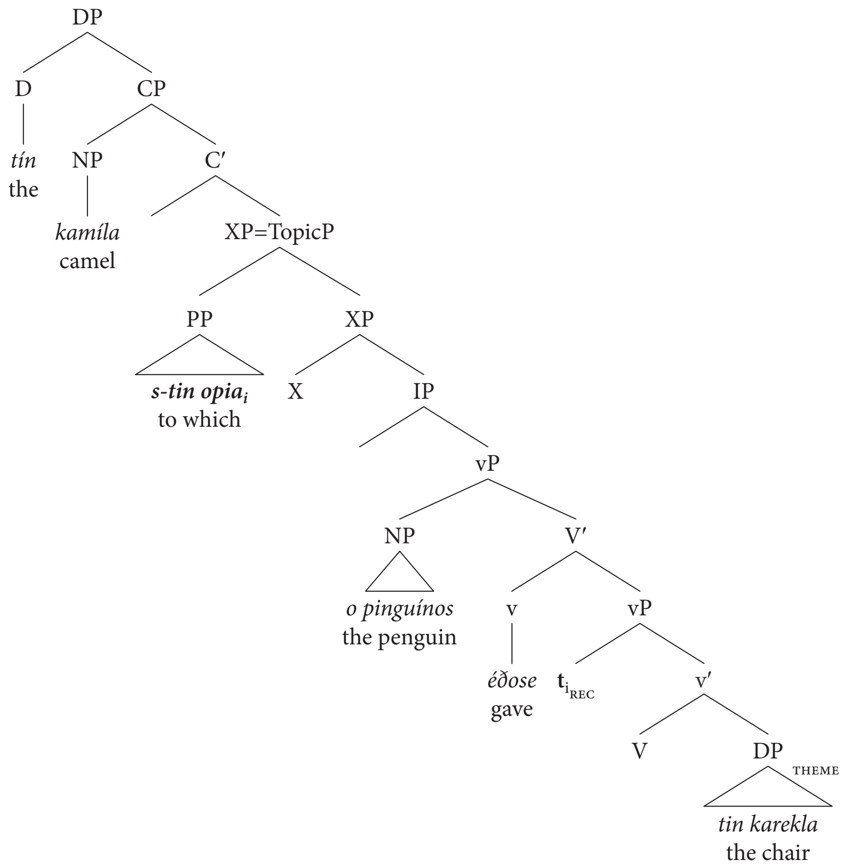

Figure 3. The base-generated position of the IO dominates the DO in an IO RC (not used due to methodological concerns)

Crucially, experimental evidence supports a base-generated account in Dutch scrambling by using the CMP paradigm. Van de Koot et al. (2016) did not find trace reactivation effects with DO scrambling over an adjunct. They concluded that their results were not suggestive of an A-scrambling analysis, noting though that these findings were restricted to the structure tested, namely the DO scrambling over a modifier and not over an argument.

4.1 Experiment 1

4.1.1 *Method*

Participants. Forty-eight adult native speakers of Greek (25 women) participated in this experiment. Their mean age was 23.5 (*SD* = 2.21, range = 18–30).

Materials. The test materials for this experiment included IO RCs, such as in (5), which are similar to those used by Miller (2015) for French, and Roberts et al. (2007) for English.

(5) o Jánis iðe tin kamila s- tin opia
The John see$_{\text{PAST}}$-3$_{\text{SG}}$ the$_{\text{ACC-F-SG}}$ camel$_{\text{ACC-F-SG}}$ to-the$_{\text{ACC-F-SG}}$ which$_{\text{ACC-F-SG}}$
o ómorfos pinguinos éðose tin kócini
the$_{\text{NOM-M-SG}}$ beautiful$_{\text{NOM-M-SG}}$ penguin$_{\text{NOM-M-SG}}$ give$_{\text{PAST-3SG}}$ the$_{\text{ACC-F-SG}}$ red$_{\text{ACC-F-SG}}$
karékla ti Dheftéra s- to párti.
chair$_{\text{ACC-F-SF}}$ the-$_{\text{ACC-F-SG}}$ Monday at-the$_{\text{ACC-N-SG}}$ party.
'John saw the camel to which the beautiful penguin gave the red chair on Monday at the party.'

The materials comprised 36 experimental and 71 filler sentences (distractors). The filler sentences included a variety of complex sentences, such as adverbial and subject RCs, as well as global and local ambiguities irrelevant to the structure under investigation. Two pictures were chosen as visual targets for each sentence, one depicting an animate object and one depicting an inanimate object. The animate pictures of the experimental sentences matched the antecedent of the RC (type of picture: identical) and the inanimate pictures were completely irrelevant of the content of the sentence (type of picture: unrelated picture). The pictures appeared in one of the following test points: (a) at a pre-gap position,[7] (b) at a gap position, and (c) at a post-gap position in the experimental sentences and at random positions in the filler sentences. The 3 (positions) × 2 (type of picture) design generated six different experimental conditions, as illustrated in (6a)–(6f).

(6) O Jánis iðe tin kamila stin opia o ómorfos pinguinos éðose…
a. *C1, pre-gap identical:*
…tin kócini [picture: CAMEL] karékla ti Dheftéra sto párti.
b. *C2, pre-gap unrelated:*
…tin kócini [picture: UMBRELLA] karékla ti Dheftéra sto párti.
c. *C3, gap identical:*
…tin kócini karékla [picture:CAMEL] ti Dheftéra sto párti.
d. *C4, gap unrelated:*
…tin kócini karékla [picture: UMBRELLA] ti Dheftéra sto párti.
e. *C5, post-gap identical:*
…tin kócini karékla ti Dheftéra [picture: CAMEL] sto párti.

7. An anonymous reviewer notes that the pre-gap control position of the picture splits a syntactic constituent. The pre-gap position of the picture was the same as in the Felser and Roberts (2007) and the Miller (2015) study for reasons of consistency. Additionally, to the best of our knowledge, most studies on filler integration in head-initial languages have posited the pre-gap (control) point of the picture (a) either one word or (b) 400–600 milliseconds before the critical gap position.

f. *C6, post-gap unrelated:*
…tin kócini karékla ti Dheftéra [picture: UMBRELLA] sto párti.
'John saw the camel to which the beautiful penguin gave the red chair on Monday at the party.'

The following ditransitive verbs were used in past tense and perfective aspect; *give, donate, offer, sell, present, send, lend, show, explain*. Each verb was used 4 times.

Pre-test. To explore how the animacy decision itself affects RTs, all participants were administered a picture selection pre-test. In this task, they saw the pictures of the PCDR and they had to categorize them as animate or inanimate. RTs from the time that the picture appeared to participants' pressing the right key on the keyboard were measured.

PCDR Procedure. Figure 4 represents the procedure of the task visually; participants had to read sentences aloud segment by segment. Each segment appeared on the display for 700ms. At certain points of the sentence (e.g. at the position of the gap) a picture appeared on the display and they had to indicate whether the picture was animate or inanimate by pressing the relevant key on the keyboard. In the end of the session, all participants took part in a short debriefing questionnaire, which also applied to all experiments.

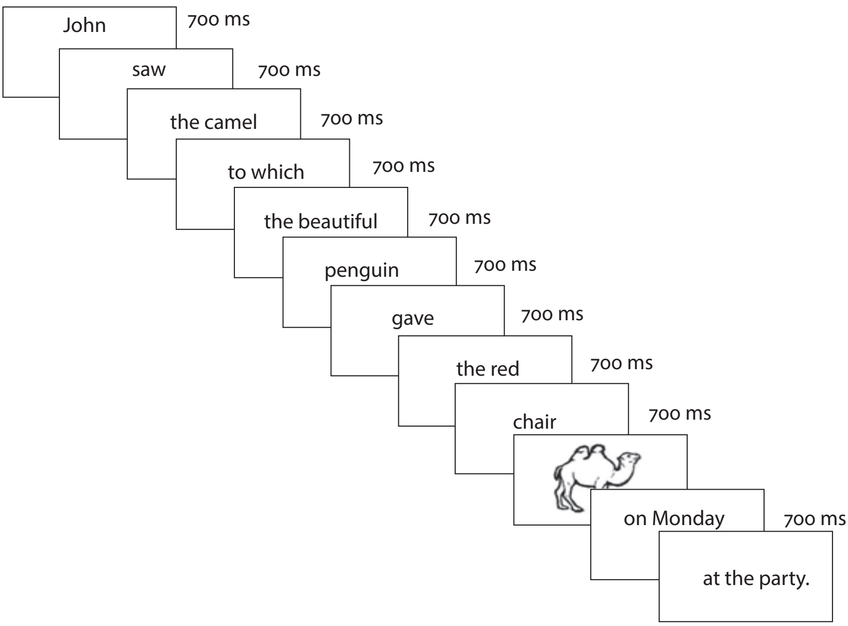

Figure 4. Procedure in Experiment 1. Here one of the experimental trials in C3 – gap, identical (see 6c) is presented

Predictions. If NSs process these sentences in line with the TRH, and if the DO–IO is the base-generated order in Greek prepositional ditransitives, we expect an interaction between position and picture (shorter RTs for the identical picture only at the gap position).

4.1.2 *Results*

The analysis of the debriefing questionnaire showed that none of the participants had become aware of the structures tested during the experiments. In the pre-test, participants performed at ceiling ($M = 96.8$, $SD = 2.36$, range = 88–99). No significant differences were detected between the means of the identical and the unrelated pictures (t_1 (47) = −1.165, $p = 0.250$, t_2 (35) = −0.570, $p = .572$) indicating that any RT difference in the picture categorization of the main task is not attributed to the categorization itself or other factors (e.g., lemma frequency).[8]

In the PCDR, the participants received high accuracy scores in the animacy decision ($M = 97.9$, $SD = 5.45$, range = 66–100). Only correct answers to animacy decision were included in the RT analyses. A cut-off point of 2,500 ms was applied to define extreme values; 8 values exceeding that point were detected and deleted, affecting 0.47% of the dataset. The data were also screened for outliers; any value which was 2 SD above or below of each participant's mean was replaced by that mean affecting a further 1.8% of the dataset. Table 1 presents RTs across the conditions.

Table 1. Mean reaction times (standard deviations) and difference between identical and unrelated targets for Experiment 1

Condition	Mean RTs	Difference
C1, pre-gap identical	780 (192)	
C2, pre-gap unrelated	757 (199)	−23
C3, gap identical	767 (195)	
C4, gap unrelated	743 (198)	−24
C5, post-gap identical	780 (218)	
C6, post-gap unrelated	766 (217)	−14

An ANOVA with the factors Position and Picture as within-subject factors showed no significant main effects or interactions (F_1 (2, 94) = 0.117, $p = .889$, F_2 (2, 70) = 0.082, $p = .922$).

8. In the following experiments, participants completed the same pre-test but any differences did not turn out significant ($p > .05$).

4.2 Experiment 2

4.2.1 *Method*

Participants. Twenty-four adult native speakers of Greek (10 women) with mean age 23.2 (*SD* = 1.91, range = 19–27) participated in this experiment.

Material. The same materials as in Experiment 1 were used (see (6)). We also created comprehension questions for all experimental sentences and half of the filler sentences (Marinis, 2003; Marinis 2010). All questions included theta-role reversal to ensure that participants pay attention and process the sentences syntactically.

Procedure. The task employed was CMP and it is described in a previous section (see Methodological Background).

Predictions. Predictions are the same as in Experiment 1.

4.2.2 *Results*

Participants' accuracy to the comprehension questions was 81.5% (*SD* = 8.84, range = 63–97) indicating that they paid attention to the task.

Table 2. Mean reaction times (standard deviation) and difference between identical and unrelated targets for Experiment 2

Condition	Mean RTs	Difference
C1, pre-gap identical	732 (163)	
C2, pre-gap unrelated	765 (166)	33
C3, gap identical	766 (171)	
C4, gap unrelated	751 (180)	−15
C5, post-gap identical	739 (154)	
C6, post-gap unrelated	730 (158)	−9

An ANOVA with the factors Position and Picture as within-subject factors showed no significant main effects or interactions (F_1 (2, 46) = 1.874, p = .165, F_2 (2, 70) = 0.944, p = .394).

4.3 Experiment 3

4.3.1 *Method*

Participants. Forty-two Greek adults (20 women) with mean age 23.3 (*SD* = 1.9, range = 18–30), participated in Experiment 3.

Materials. In this experiment, as in the following one, DO RCs were used and the IO-DO was adopted as basic. As a result, experimental sentences were modified from IO RCs to DO RCs with slight changes in the material, as (7) shows.

(7) DO RC
O Jánis íðe tin karékla tin opía o ómorfos piŋguínos éðose sti meγáli kamíla ti Dheftéra sto párti.
'John saw the chair which the beautiful penguin gave to the big camel on Monday at the party.'

The testing points for the picture were kept the same as in the previous experiments. However, in this experiment the identical picture of the RC is also inanimate like the unrelated picture, as in (8). Thus, most of the filler sentences included only animate pictures for balance.

(8) O Jánis iðe tin karékla tin opia o ómorfos pinguinos éðose…
  a. *C1, pre-gap identical:*
     …sti meγáli [picture: CHAIR] kamila ti Dheftéra sto párti.
  b. *C2, pre-gap unrelated:*
     …sti meγáli [picture: UMBRELLA] kamila ti Dheftéra sto párti.
  c. *C3, gap identical:*
     …sti meγáli kamila [picture: CHAIR] ti Dheftéra sto párti.
  d. *C4, gap unrelated:*
     …sti meγáli kamila [picture: UMBRELLA] ti Dheftéra sto párti.
  e. *C5, post-gap identical:*
     …sti meγáli kamila ti Dheftéra [picture: CHAIR] sto párti.
  f. *C6, post-gap unrelated:*
     …sti meγáli kamila ti Dheftéra [picture: UMBRELLA] sto párti.
  'John saw the chair which the beautiful penguin gave to the big camel yesterday at the party. '

Procedure. A CMP task was conducted as in Experiment 2.

Predictions. If NSs process these sentences in line with the TRH, and if the IO-DO is the base-generated order in Greek prepositional ditransitives, we expect an interaction between position and picture.

4.3.2 *Results*

Participants' accuracy to comprehension questions was 83.6% (*SD* = 5.86, range = 72–94. With respect to animacy, participants were 98.7% (*SD* = 2.29, range = 92–100) correct. Data trimming applying the same criteria as previously affected 3.31% of the data. Table 3 summarizes mean RTs across the conditions.

Table 3. Mean RTs (standard deviation) and difference between identical and unrelated targets for Experiment 3. Statistically significant differences are marked with an asterisk

Condition	Mean RTs	Difference
C1, pre-gap identical	846 (138)	
C2, pre-gap unrelated	832 (148)	−14
C3, gap identical	758 (151)	
C4, gap unrelated	816 (136)	**58***
C5, post-gap identical	831 (134)	
C6, post-gap unrelated	822 (129)	−9

The ANOVA revealed a significant interaction between Position and Picture in the subject analysis (F_1 (2, 82) = 5.251, p = .007) and a marginal interaction in the item analysis (F_2 (2, 70) = 2.834, p = .066). Pairwise comparisons revealed a priming effect at the gap position, where the RTs for the identical picture (C3) (M = 758, SD = 151) were significantly shorter compared to the unrelated picture (C4) (M = 816, SD = 136) (t_1 (41) = −2.668, p = .011, t_2 (35) = −1.950, p = .059). The identical picture at the gap (C3) was also significantly shorter compared to the pre-gap identical picture (C1) (M = 846, SD = 138) (t_1 (41) = 3.172, p = .003, t_2 (35) = 3.250, p = .003) and the post-gap identical picture (C5) (M = 831, SD = 134) (t_1 (41) = −3.315, p = .002, t_2 (35) = −2.719, p = .010). These findings are indicative of trace reactivation. All other pairwise comparisons were not significant.

4.4 Experiment 4

4.4.1 *Method*

Participants. Thirty-six Greek adults (27 women) with a mean age of 21.3 (SD = 1.86, range = 18–26) participated in Experiment 4.

Materials. The same experimental materials were used as in Experiment 3.

Procedure. This task was a PCDR, as in Experiment 1.

Predictions. If the PCDR can capture the trace reactivation effect found in CMP (Experiment 3) where IO–DO was considered to be the base-generated order in Greek prepositional ditransitives, an interaction between position and picture is expected.

4.4.2 *Results*

Participants scored at ceiling (97.2%, *SD* = 2.89, range = 91–100) in the picture selection. The same data trimming criteria were applied, affecting 0.5% of the dataset. Table 4 summarizes mean RTs across the conditions.

Table 4. Mean RTs (standard deviation) and difference between identical and unrelated targets for Experiment 3

Condition	Mean RTs	Difference
C1, pre-gap identical	862 (202)	
C2, pre-gap unrelated	843 (247)	−19
C3, gap identical	841 (230)	
C4, gap unrelated	818 (218)	−23
C5, post-gap identical	825 (197)	
C6, post-gap unrelated	840 (248)	15

The ANOVA did not detect any significant main effects or interactions between Position and Picture (F_1 (2, 70) = 0.802, p = .453, F_2 (2, 70) = 0.790, p = .458).

5. Discussion

The present study addressed filler integration in object RCs in adult native speakers of Greek using four experiments. This enabled us to test two different theoretical accounts for the base word order and the hierarchical position of the arguments as well as the effectiveness of two experimental paradigms, CMP and PCDR. In Experiment 1 we addressed filler integration using IO RCs adopting the theoretical account which proposes that the DO is posited at a higher hierarchical position than the IO and employed a PCDR task. The results did not reveal any processing pattern. Thus, these results do not confirm the TRH in Experiment 1. However, the lack of significant effects could be explained by the DAH or any other trace-free account e.g., the Good Enough approach (Ferreira & Patson, 2007), in the sense that the accounts that dispute the psychological reality of traces would not predict any effect at the gap position.

We then investigated two alternatives, which subsequently generated Experiments 2 and 3. To explore whether the findings of Experiment 1 are attributed to the task, we conducted Experiment 2 applying the same experimental material in a CMP task. The findings of Experiment 2 are in line with Experiment 1. This indicates that the results are not the outcome of the specific procedure used. In Experiment 3 we adopted a different theoretical account on the hierarchy of ditransitive arguments, namely that the IO is posited at a higher hierarchical position than the DO, hence the IO–DO is considered to be base-generated. To empirically dissociate the verb from the gap, DO RCs were used where the IO spans between the verb and the gap. Experiment 3 used a CMP task and revealed trace reactivation effects. This confirmed both the TRH account and the psychological reality of the IO–DO order as basic.

To address task effects between CMP and PCDR, we conducted Experiment 4 using the material from Experiment 3 in a PCDR task. The findings from Experiment 3 were not replicated in Experiment 4. This suggests that the PCDR paradigm is less sensitive than the CMP paradigm to test filler integration.

One of the shortcomings of this study is that different structures were employed to test the two syntactic analyses, IO RCs and DO RCs. We are aware of the fact that there might be an asymmetry in comprehension (and perhaps processing load) between IO RCs and DO RCs. For example, it could be claimed that DO RCs are easier to comprehend because of the tight relationship between the DO and its verb (Georgiafentis & Lascaratou, 2007). However, participants' accuracy to comprehension questions in Experiments 2 (81.5% in IO RCs) does not differ from participants' accuracy in Experiment 3 (84.5% in DO RCs). This indicates that participants were able to comprehend both structures equally well.

Furthermore, the current findings are different from two other recent studies (Sekerina et al., 2018; van de Koot et al., 2016) testing trace reactivation, which may be attributed to the different structures tested. In the current experiments, the filler as well as the position of the trace were always referring to an argument of the verb, i.e., DO and IO. Van de Koot et al. tested trace reactivation of the DO over an adjunct and not over another argument (e.g. IO) as we did in the current study. Based on the absence of reactivation effects, the authors concluded that A-scrambling is not movement in Dutch and that it results from variation in base-generated order. Similarly, Sekerina et al. also tested object scrambling in wh-questions in Russian and they did not find trace reactivation. Thus, the different types of movement and/or the different structures tested in these studies may be responsible for the lack of trace reactivation.

Future work can shed more light into the trace reactivation and its consequences for syntactic theories as well as for theories of language comprehension. For example, certain structures can be selected and compared both at the level of a certain language as well as cross-linguistically. A comprehensive study of potential trace reactivation in structures under A vs A′ movement across different languages would be informative about the similarities and differences between these two types of movements during online processing.

The current study also showed that CMP was sensitive to the phenomenon tested, while PCDR was not. The lack of any effect with PCDR is possibly related to the fact that this task does not include comprehension questions in the end of the trials, and consequently, it cannot verify participants' attention while performing the task. Thus, it seems that reading aloud does not necessarily guarantee that participants parse the sentence syntactically.

To conclude, this study revealed trace reactivation effects in Greek DO relative clauses using CMP. The reactivation was in line with the IO–DO analysis, according to which IO are higher hierarchically than DO, thus, providing evidence for the psychological reality of this account. Finally, the CMP paradigm was shown to be more sensitive for the detection of trace reactivation in adult native speakers of Greek than the PCDR.

Aknowledgements

I am grateful to Theodoros Marinis, Despina Papadopoulou and the audience of the 37th Annual Meeting of the Department of Linguistics – AUTH for helpful comments and discussion. An earlier version of this work has been published in the Proceedings of the 37th Annual Meeting of the Department of Linguistics – AUTH.

Funding

This work was partially funded from AL 554/8-1, DFG Gottfried Wilhelm Leibniz Preis 2014 awarded to Artemis Alexiadou.

References

Anagnostopoulou, E. (2003). Participles and voice. In A. Alexiadou, M. Rathert, & A. von Stechow (Eds.), *Perfect explorations* (pp. 1–36). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110902358.1

Anagnostopoulou, E. (2005). Cross-linguistic and cross-categorical variation of datives. In M. Stavrou & A. Terzi (Eds.), *Advances in Greek generative syntax. In Honor of Dimitra Theophanopoulou-Kontou* (pp. 61–126). Amsterdam: John Benjamins. https://doi.org/10.1075/la.76.05ana

Anagnostopoulou, E. (2013). Greek Syntax 2012: State-of-the art and perspectives. In M. Stavrou & S. Michalopoulou (Eds.), *Studies in Greek linguistics. Proceedings of the 33rd Annual Meeting of the Department of Linguistics, Aristotle University of Thessaloniki* (pp. 11–58). Thessaloniki: Institute of Modern Greek Studies (Manolis Triantaphyllidis Foundation).

Barss, A., & Lasnik, H. (1986). A note on anaphora and double objects. *Linguistic Inquiry*, 17, 347–354.

Bowers, J., & Georgala, E. (2007). The syntax of goals and beneficiaries in Modern Greek. In A. Alexiadou (Eds.), *Studies in the morpho-syntax of Greek* (pp. 14–46). Newcastle upon Tyne: Cambridge Scholars.

Clahsen, H., & Featherston, S. (1999). Antecedent-priming at trace positions: Evidence from German scrambling. *Journal of Psycholinguistic Research*, 28, 3415–437. https://doi.org/10.1023/A:1023293132656

Dekydtspotter, L., & Miller, K. (2009). Probing for intermediate traces in the processing of long-distance wh-dependencies in English as a second language. In M. Bowles, T. Ionin, S. Montrul, & A. Trembley (Eds.), *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference* (pp. 113–124). Somerville, MA: Cascadilla Press.

Dekydtspotter, L., & Miller, K. (2013). Inhibitive and facilitative priming induced by traces in the processing of wh-dependencies in a second language. *Second Language Research*, 29, 345–372. https://doi.org/10.1177/0267658312467030

Dekydtspotter, L., Miller, K., Schaefer, V., Chang, Y., & Kim, O. H. (2010). Clause-edge reactivation of fillers in processing English as a second language. In M. T. Prior, Y. Watanabe, & S. Lee (Eds.), *Selected Proceedings of the 2008 Second Language Research Forum* (pp. 108–122). Somerville, MA: Cascadilla Press.

Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research*, 23, 9–36. https://doi.org/10.1177/0267658307071600

Ferreira, F., & Patson, N. (2007). The 'Good Enough' approach to language comprehension. *Language and Linguistics Compass*, 1, 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Georgala, E. (2012). *Applicatives in their Structural and Thematic Function. A minimalist account of Multitransitivity* (Unpublished doctoral dissertation). Cornell University.

Georgala, E., & Whitman, J. (2009). Ditransitives and applicative structure in Greek. In N. Adams, A. Cooper, F. Parrill, & T. Wier (Eds.), *Proceedings from the Annual Meeting of the Chicago Linguistic Society* (pp. 77–91). Chicago, IL: CLS.

Georgiafentis, M., & Lascaratou, C. (2007). On the inherently cloze relation of the verb and its direct object in ditransitive constructions: Evidence from Greek. *Journal of Greek Linguistics*, 8, 31–59. https://doi.org/10.1075/jgl.8.05geo

Holton, D., Mackridge, P., & Philippaki-Warburton, I. (1997). *A comprehensive grammar of the Modern Greek language*. London: Routledge.

Marantz, A. (1993). Implications of asymmetries in double object construction. In S. Mchombo (Eds.), *Theoretical aspects of Bantu grammar* (pp. 113–150). Stanford, CA: CSLI.

Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Second Language Research*, 19, 144–161. https://doi.org/10.1191/0267658303sr217ra

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In S. Unsworth & E. Blom (Eds.), *Experimental methods in language acquisition research* (pp. 139–162). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.27.09mar

Marinis, T. (2018). Crossmodal priming in bilingual sentence processing. *Bilingualism: Language and Cognition*, 21(3), 456–461. https://doi.org/10.1017/S1366728917000761

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, 27, 53–78. https://doi.org/10.1017/S0272263105050035

Michelioudakis, D. (2011). *Dative arguments and abstract case in Greek* (Unpublished doctoral dissertation). University of Cambridge.

Miller, K. (2011). Rethinking nonnative processing constraints: Evidence from L2 French. In L. Plonsky & M. Schierloh (Eds.), *Selected proceedings of the 2009 Second Language Research Forum* (pp. 109–120). Somerville, MA: Cascadilla Press.

Miller, K. (2013). Second language processing of filler-gap dependencies: Evidence for the role of lexical access. In E. Voss, S. J. Tai, & Z. Li (Eds.), *Selected proceedings of the 2011 Second Language Research Forum* (pp. 82–94). Somerville, MA: Cascadilla Press.

Miller, K. (2015). Facilitating the task for second language processing research: A comparison of two testing paradigms. *Applied Psycholinguistics*, 36, 613–637. https://doi.org/10.1017/S0142716413000362

Nakano, Y., Felser, C., & Clahsen, H. (2002). Antecedent priming at trace positions in Japanese long-distance scrambling. *Journal of Psycholinguistic Research*, 31, 531–571. https://doi.org/10.1023/A:1021260920232

Nicol, J. (1993). Reconsidering reactivation. In G. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 321–350). Hove: Psychology Press.

Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognition Processes*, 6, 229–259. https://doi.org/10.1080/01690969108406944

Pylkkänen, L. (2002). *Introducing Arguments.* (Unpublished doctoral dissertation). MIT.

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistic Research*, 36, 175–188. https://doi.org/10.1007/s10936-006-9038-3

Sekerina, I., Laurinavichyute, A., & Dragoy, O. (2018). What eye movements can and cannot tell us about wh-movement and scrambling. In K. Carlson, C. Clifton, Jr., & J. D. Fodor (Eds.), *Grammatical approaches to language essays in honor of Lyn Frazier* (pp. 147–165). Berlin: Springer.

Swinney, D., Onifer, W., Prather, P. & Hirshkowitz, M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory and Cognition*, 7, 54–178. https://doi.org/10.3758/BF03197534

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35, 454–475. https://doi.org/10.1006/jmla.1996.0025

Van de Koot, H., Silva, R., Felser, C. & Sato, M. (2016). Does Dutch a-scrambling involve movement? Evidence from antecedent priming. *The Linguistic Review*, 34, 739–776.

When initial thematic role attribution lingers

Evidence for digging-in effects in Italian relative clauses

Sandra Villata[1] and Paolo Lorusso[2,3]
[1]University of Connecticut / [2]IUSS Scuola Universitaria Superiore Pavia / [3]CRIL Università del Salento

In Italian, relative clauses are syntactically ambiguous between a subject and an object reading when the subject and the object have the same number. In the absence of disambiguating cues, the parser analyzes the sentence as a subject relative clause, as subject relative clauses are easier to process than object relative clauses. However, the object reading can be triggered by morpho-syntactic and syntactic cues, such as agreement and word order. In two self-paced reading experiments, we investigate the effectiveness of these cues in triggering an object reanalysis by comparing cues that appear at an *early* processing stage (inside the relative clause) and cues that appear at a *late* processing stage (after the relative clause). Results show that reanalysis is attained more effortlessly for early cues, while late cues appears to trigger little or no reanalysis, suggesting that the longer the processor has been committed to the incorrect analysis, the harder is to undo the initial commitment (*digging-in effect*; Tabor & Hutchins, 2004; see also Ferreira & Henderson, 1991). From these results, we argue that one critical factor that affects ease of parsing is the processing stage at which a cue is made available in the input. We conclude with a discussion on the self-organized sentence processing model (SOSP), which provides a framework to capture these effects.

Keywords: Italian, relative clauses, ambiguity, reanalysis, digging-in effect, self-paced reading, self-organized sentence processing model, SOSP, cue-based memory model

https://doi.org/10.1075/lald.64.03vil

1. Introduction

Italian relative clauses (RCs, henceforth) can be ambiguous between a subject and an object interpretation: when the two DPs (subject and object) have the same number marking, the second DP (i.e. *il pagliaccio* – "the clown" in (1)) can be interpreted either as the object of the verb *disegnare* ("to draw") or as its subject, since Italian permits subjects to appear post verbally. This is illustrated in (1), which is ambiguous between a subject (2) and an object RC reading (3).

(1) La bambina che disegna il pagliaccio ride
The girl that draws the clown smiles

(2) Subject reading
$[_{IP}[_{DP}$La bambina$_{iCP}$[che$_{i\ IP}$[*pro*$_i$[$_I$disegna [$_{\nu P}$t$_i$[$_{VP}$…[$_{DP}$il pagliaccio$_j$]]]]]]]
[$_I$ride$_i$…]]
The girl who draws the clown smiles

(3) Object reading
$[_{IP}[_{DP}$La bambina$_{iCP}$[che$_{i\ IP}$[*pro*$_j$[$_I$disegna [$_{\nu P}$ il pagliaccio$_j$[$_{VP}$.[$_{DP}$t$_i$]]]]]]]
[$_I$ride$_i$…]]
The girl that the clown draws smiles

A very robust cross linguistic finding is that object RCs are harder to process than subject RCs in language acquisition (Frauenfelder et al. 1980, Friedmann et al. 2009, Adani et al. 2010, Belletti et al. 2012 among many others), in adult processing (De Vincenzi 1991; King and Kutas 1995, Schriefers et al. 1995, Mak, Bonk & Schriefers 2002, 2006 among many others) and in pathological populations (e.g. Caramazza & Zurif 1976; Garraffa & Grillo, 2008; Grillo 2009). When confronted with a globally ambiguous sentence like (1), the parser prefers the subject RC analysis (2), in line with the vast literature showing that the parser engages with the analysis that reduces processing demands (e.g. Frazier 1978, 1987; Frazier & Rayner 1982; Minimal Chain Principle, De Vincenzi 1991). The object reading, however, can be triggered by two grammatical cues: i) word order (OSV, e.g., *The girl that the clown draws smiles*), and ii) number agreement (the verb of the RC agrees in number with the second noun phrase, thus signaling that the subject occupies the post verbal position, e.g., *The girl-SG that draw-PL the clowns-PL smiles*) (e.g., Arosio, Adani, & Guasti, 2009; Guasti, Vernice, & Franck, 2018).

In two self-paced reading studies conducted with Italian-speaking and French-speaking adults, Guasti et al., 2018 found that, in both languages, (i) participants were slower in processing object RCs as compared to subject RCs at the relative clause region, and (ii) object RC disambiguated by number agreement were read slower than object RC disambiguated by word order at the relative clause region.

From these findings, the authors concluded that word order is a stronger disambiguating cue than number. Similar results were obtained by Arosio et al. (2009) for Italian-speaking adults using a picture selection task: participants were more accurate for object RCs disambiguated by word order than for object RCs disambiguated by number agreement.

Although these studies provide an indication for an advantage of word order as compared to number as a disambiguating cue, they come with important shortcomings. Specifically, Guasti et al. concluded that the object reading interpretation was attained on the basis of a slowdown in reading times in the two object relative conditions as compared to the subject relative condition. However, since no comprehension accuracy measure targeting the thematic role attribution in the RC was collected, these results do not inform us as to whether participants actually revised their initial interpretation. A recent study conducted on Italian adult speakers showed that participants display poor comprehension of object relative clauses, with up to 35% errors (Villata, Tabor, & Franck, 2018). Given this extremely high percentage of errors, it is essential to collect direct measures that inform us if the correct parse was ultimately built.

As far as Arosio et al. study is concerned, their results are based on a picture matching task, which is an extremely coarse-grained procedure for adults, as attested by the almost at ceiling accuracy they obtained, which is at odds with the results of Villata et al. 2018. In Experiment 1, we thus investigate whether word order and number agreement cues successfully trigger the object reading in Italian RCs. With this aim, we employ a self-paced reading procedure combined with a sentence comprehension task targeting thematic role attribution inside the RC, which provides us with direct evidence about the parse that was ultimately built.

A second limitation of the studies by Guasti et al., 2018 and Arosio et al., 2009 concerns the fact that the disambiguating cue was always presented at an early processing stage, i.e. inside the RC. Models of sentence processing make different predictions as to whether the processing stage at which a cue is made available in the input influences its effectiveness in disambiguating the sentence. According to the cue-based memory model (e.g. Lewis & Vasishth, 2005; McElree, 2000; McElree et al., 2003) the answer is negative. In a cue-based memory architecture, a cue is effective if it is unambiguous, i.e. if it is compatible with one and only one analysis. For instance, there is consistent evidence in the literature showing that the retrieval from memory of a previously encountered element is impaired if the retrieval cue matches multiple elements in memory. In a memory load study, Van Dyke & McElree (2006) showed that participants were slower in reading the verb of an object cleft (e.g. *It was the boat that the guy who lived by the sea* ***sailed/fixed*** *in two sunny days*) when the semantic cues on the verb matched both the semantic feature of the object of the verb (*the boat*) and those of distracting elements in the memory load (*table, sink, truck*) (the fixable condition), than when they only matched the features of the object (the sailable condition). These results suggest that retrieval is harder when the retrieval cues are not unique to the target but also match other irrelevant elements in memory. In this model, the probability to build a correct parse is thus a function of the degree of match between the cue and the target, but also of the resonance between the cue and other irrelevant elements in memory. Importantly, the timing at which a cue is made available in the input is irrelevant in this architecture: all that matters is the accuracy with which a given cue disambiguates the input.

A different standpoint is taken by the self-organized sentence processing model (SOSP) (e.g. Tabor & Hutchins, 2004, Smith & Tabor, 2018; Villata, Sprouse, & Tabor, 2019). Tabor and Hutchins (2004) tested sentences such as *As the author wrote (the essay) the book grew* and *As the author wrote (the essay) the book describing Babylon grew*, where both the transitivity of the sentence (transitive vs. intransitive) and the length of the ambiguous region (short vs. long) were manipulated. They observed lower comprehension accuracy and longer reading times at the verb *grew* in the intransitive long condition as compared to the intransitive short condition. They interpreted this result as follows: the erroneous linkage between *wrote* and *the book* grows stronger when the ambiguous region is long than when it is short, thus becoming harder to undo. This is the result of the passage of time, but also of the formation of additional consistent linkages that strengthen the initial analysis. In other words, the longer the parser has been committed to one analysis, the harder the reanalysis is expected to be, an effect known as *digging-in* (Tabor & Hutchins, 2004; see also Ferreira & Henderson, 1991). Under SOSP, digging-in effects are the result of competition for attachment during structure building. In this model, lexical items carry syntactic and semantic requirements that inform the formation of the attachments during structure building. Attachments between lexical items with a good feature match generally outcompete attachments between elements with a poor feature match, which ensures that, most of the time, a well-formed structure is generated. Attachments between lexical elements grow in strength as a result of the passage of time and additional consistent linkages. This property of the systems is at the source of digging-in effects. SOSP thus predicts that the moment at which a cue is made available in the input influences structure building operations: late cues are expected to be less effective than early cues in disambiguating the sentence.

In order to tease apart predictions from these two models, Experiment 2 investigates the effectiveness of a *late* gender agreement cue appearing after the RC in revising the initial subject reading into an object reading. If the moment at which a cue is made available in the input is a factor influencing the ease with which the parser reanalyzes the sentence, in line with predictions from SOSP, then late cues are expected to be less effective than early cues in triggering reanalysis. On the contrary, if when a cue is made available in the input is irrelevant for structure building operations, and all what matters is the precision with which a cue signals the correct analysis, in line with predictions from the cue-based memory model, then late cues are expected to be as effective as early cues.

Summarizing, in the next two sections we report two self-paced reading studies combined with a sentence comprehension task in which we tested, respectively: (i) the disambiguating effectiveness of number and word order cues that appear early in the sentence (i.e. inside the RC) in triggering the object reading of ambiguous Italian RCs (Section 2), and (ii) the effectiveness of an agreement cue appearing late in the sentence (i.e. after the RC) in triggering object reanalysis (Section 3).

2. Experiment 1: Disambiguation by early cues

2.1 Method

2.1.1 *Participants, materials and design*

Sixty-seven Italian native speakers participated in this experiment (46 female and 21 male, age: 21–63, mean age: 34). We created 32 sets of three conditions each illustrated in (4). We manipulated the *Type of RC* (subject relative, SR vs. object relative, OR), and the *Type of disambiguation cue* (word order vs. number agreement).

(4) a. **SVO subject relative clause (number cue)**
I calciatori-PL/ che/ chiamano-PL/ l'arbitro-SG/ nel/ secondo/
The footballers who call the referee during the second half
tempo /hanno-PL /vinto /i Mondiali.
have won the World Cup

b. **OSV object relative clause (word order cue)**
Il calciatore-SG/ che/ l'arbitro-SG/ chiama-SG/ nel/ secondo/
The footballer who the referee calls during the second
tempo/ ha-SG/ vinto/ i Mondiali.
half has won the World Cup

c. **OVS object relative clause (number cue)**
I calciatori-PL/ che /chiama-SG/ l'arbitro-SG/ nel/ secondo/
The footballers who calls the referee during the second
tempo/ hanno-PL/ vinto/ i Mondiali.
half have won the World Cup

In the subject relative clause condition (4a), the first DP was always plural while the second DP was singular. Since the relative verb (e.g., *chiamano-PL*) agreed in number with the first DP, the sentence was unambiguously a subject relative clause. In OSV object relative clauses (4b), both DPs were singular, but they both occurred pre-verbally, therefore triggering an object relative clause reading. In OVS object relative clauses (4c), the first DP was always plural, while the second DP was always singular and the relative verb (e.g., *chiama-SG*) agreed in number with the second DP, thus triggering an object relative clause reading. The matrix past participle verb (e.g., *hanno/ha vinto*) was always introduced by a temporal modifier and followed by a direct object. Experimental sentences were split in 8 lists in order to reduce the number of experimental sentence participants were confronted with. Experimental sentences were intermixed with 32 fillers, which consisted of 12 relative clauses (including cleft sentences and free relatives), which served as experimental sentences for another experiment not reported here, and 20 complex sentences not containing any RC. In total, each participant read 64 sentences (32 experimental items and 32 fillers).

2.1.2 *Procedure*

The experiment was programmed with the E-prime software (Schneider, Eschman, and Zuccolotto, 2012). We used a non-cumulative self-paced paradigm procedure (Just et al., 1982) where participants were instructed to read the sentences word-by-word by pressing the space bar in order to have the segments appear. Each segment was presented in the middle of the screen and disappeared as soon as the participant pressed the space bar. Participants were told that a yes/no comprehension question would be asked at the end of each sentence. Comprehension questions specifically targeted thematic role attribution inside the RC (e.g., *Did the footballer call the referee? vs. Did the referee call the footballer?*) in order to determine whether the correct parse was built. Half the questions required a 'yes' answer. Items were presented in a fixed pseudo-random order, constrained such that experimental sentences were always separated by one filler. Each experimental session began with six practice trials. The whole session lasted about 20 minutes.

2.1.3 *Data analyses*

Reading times were analyzed by way of linear mixed-effects regression models, and for question answering data we used generalized linear mixed-effects regression models using the lme4 package (Bates et al., 2015) in the R software environment (R Development Core Team, 2016). Reading times greater than 3000 ms or less than 100 ms were removed (affecting 2% of the data). Reading times were log-transformed to normalize residuals and then regressed against two factors that may affect reading times, namely word length and the log list position of the sentence in the stimuli. The residual log reading time is therefore the dependent variable analyzed here. Error bars in graphs represent standard errors by subject means. Because we are interested in investigating the relative advantage of word order and agreement cues in driving structure building, analyses on reading times were conducted on correct trials only, i.e. trials for which participants built the correct parse. All the predictive factors were dichotomous and centered by coding one level of the factor as −1 and the other as 1. We always used the maximal random-effects structure by participant and by item justified by the data. No correlations between random effects were estimated. P-values were calculated by way of the Satterthwaites's approximation to degrees of freedom with the lmerTest package (Kuznetsova et al., 2016). To assess the relative clause effect, we analyzed the region in which thematic roles are integrated with the relative verb across conditions (region 4).

2.1.4 *Results*

Reading times. Figure 1 illustrates the distribution of reading times across the three experimental conditions. We plotted non-transformed reading times for readability, but analyses were conducted on residual log reading times.

Region 4. Results revealed that subject relatives were read significantly faster (M = 887 ms) than both the object relative OSV condition (M = 1073; $\beta = 0.2424$, $SE = 0.049$, $t = 4.970$, $p < 0.001$) and the object relative OVS condition (M = 1125 ms, $\beta = 0.267$, $SE = 0.061$, $t = 4.412$, $p < 0.001$). No difference was attested between the two object relative clause conditions ($t < 1$).

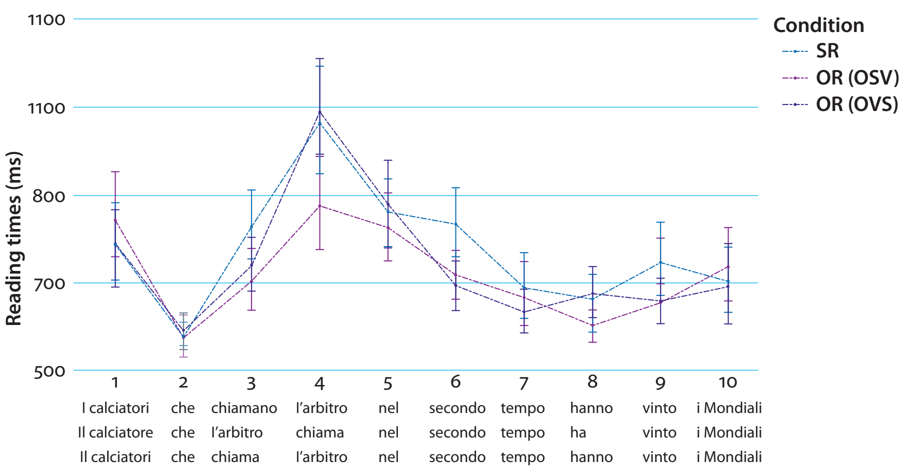

Figure 1. Distribution of reading times (in ms) in the three experimental conditions for the different regions in Experiment 1 (correct trials only)

Comprehension accuracy. Figure 2 illustrates the distribution of accuracy proportions for the three experimental conditions. Results showed a main effect of the type of relative clause ($\beta = 5.343$, SE = 1.565, $z = 3.413$, $p < .001$), with higher accuracy proportions for subject relative clauses ($M = 0.88$) than object relative clauses ($M = 0.76$). In order to investigate the effect of the type of cue in triggering the object reading, we ran an additional model on object relative clauses only. No significant effect was found ($z < 1$).

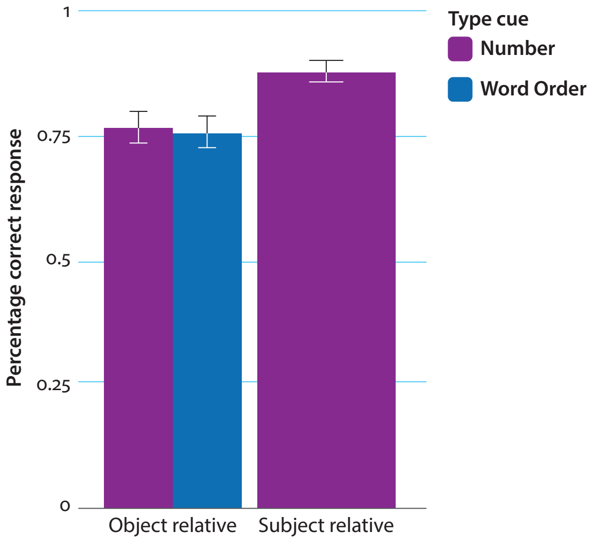

Figure 2. Accuracy proportion in Experiment 1

2.1.5 *Discussion*

Experiment 1 found an advantage of subject RCs as compared to object RCs, such that subject RCs were read faster at the integration region (region 4) and were better understood than object RCs, in line with previous researches (e.g., De Vincenzi, 1991; Belletti & Contemori, 2010; Guasti et al., 2018 among many others). Critically, no difference was attested between object RCs disambiguated by number and those disambiguated by word order, neither online nor in comprehension accuracy results. Our online results are thus in contrast with results by Guasti et al. 2018, who reported longer reading times in adults for OVS than for OSV object relative clauses, thus concluding that word order was a stronger disambiguating cue than number agreement in Italian. However, while Guasti et al., 2018 collapsed the complementizer, the RC verb region and the second DP region in reading times analyses, we kept these regions separated, a difference that might be responsible for the difference between our results and Guasti et al.'s results. We therefore conducted an additional analysis collapsing the three regions (i.e. *che* *chiamano l'arbitro*), but no difference between the two OR clauses was attested ($t < 1$). It is possible that Guasti et al.'s result might be a Type II error due to lack of power since the authors only tested 28 participants.

Finally, Guasti et al.'s results are silent about the final interpretation of the sentence reached by participants, thus providing no direct indication of the relative effectiveness of word order and number agreement in triggering the object reading. Our results suggest that number agreement and word order are equally effective in triggering an object reading in Italian, as we found no difference in comprehension accuracy between OVS and OSV RCs.

All in all, results from Experiment 1 revealed that both word order and number agreement are effective in triggering an object reading in Italian RCs and this to a similar extent (76% of correct responses). However, both cues appeared at early processing stages (i.e. inside the relative clause). In Experiment 2, we investigate whether the object reading can also be triggered by a cue appearing later in the sentence, which will allow us to disentangle between predictions from the cue-based memory model and SOSP.

3. Experiment 2: Disambiguation by a late cue

3.1 Method

3.1.1 *Participants, materials and design*

Thirty-four Italian native speakers participated in this experiment (18 female and 16 male, age: 19–39, mean age: 29). With the aim to test the effectiveness of a late cue appearing after the relative clause in triggering an object RC reading in Italian relative clauses, we manipulated the gender agreement of the past participle following the relative clause (*ascoltato*); see (5). The past participle agreed either with the pre-verbal DP (*il sindaco-M*) (5a), which supports the default subject RC interpretation, or with the post-verbal DP (*la giornalista-F*) (5b), which should trigger an object RC reanalysis. The two DPs always mismatch in their gender feature, such that the past participle unambiguously agreed with only one of the two DPs. All DPs were singular. Thirty-two set of sentences were created and split in 4 lists to reduce the number of relative clauses participants were confronted with. Experimental items were intermixed with 40 fillers, which were all complex sentences involving movement but not containing a relative clause. In total, each participant read 72 sentences (32 experimental items and 40 fillers).

(5) a. **SVO subject relative clause**
Il sindaco-M / che / consulta / la giornalista-F / prima / di / essere /
The mayor-M that consults the journalist-F before of being
ascoltato-M/ *da tutti / vive / a Parigi.*
heard-M by everyone lives in Paris

b. **SVO object relative clauses (late gender cue)**
Il sindaco-M / che / consulta / la giornalista-F / prima di / essere /
The mayor-M that consults the journalist-F before of being
ascoltata-F */ da tutti / vive / a Parigi.*
heard-F by everyone lives in Paris

3.1.2 *Procedure*

The procedure was the same as Experiment 1.

3.1.3 *Data analyses*

The same analyses conducted for Experiment 1 were used here, except that in this case analyses on reading times were conducted on all trials due to the low proportion of correct responses in the object RC condition. To assess the effect of the gender cue in triggering the object reading we analyzed two separate regions: the critical region containing the past participle (region 7), and the region immediately following it (region 8), in order to assess for spillover effects.

3.1.4 *Results*

Reading times. The distribution of reading times is illustrated in Figure 3. We plotted non-transformed reading times for readability, but analyses were conducted on residual log reading times.

Region 7. No effect was significant ($t < 1$).

Region 8. Result attested to a significant main effect of gender agreement, with faster reading times when the past participle agreed with the first noun phrase (subject relative clause condition, $M = 617$ ms) than when it agreed with the second noun phrase (object relative clause condition, $M = 680$ ms) ($\beta = -0.286$, $SE = 0.013$, $t = -2.132$, $p = 0.041$).

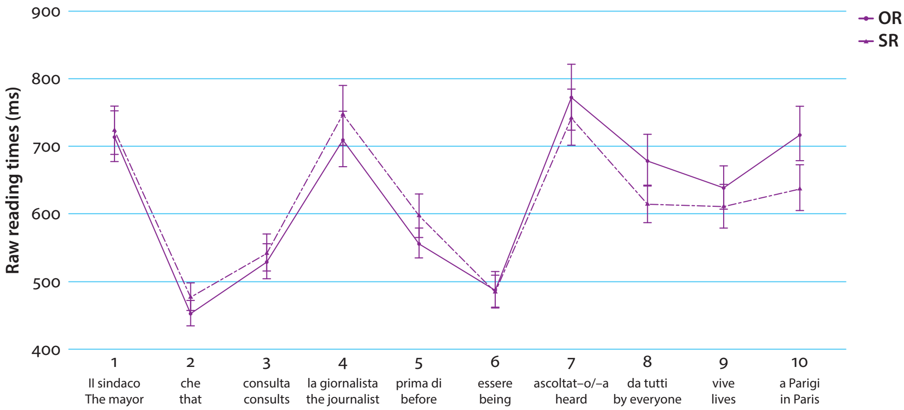

Figure 3. Distribution of reading times (in ms) in the two experimental conditions for the different regions in Experiment 2 (all trials)

Comprehension accuracy. Figure 4 illustrates the distribution of accuracy proportions in the two experimental conditions. Results showed a main effect of gender agreement, attesting to higher comprehension accuracy rates when the past participle agreed with the first noun phrase ($M = 0.83$), i.e. subject relative condition, than when it agreed with the second noun phrase, i.e. object relative condition ($M = 0.20$; $\beta = 1.613$, SE $= 0.027$, $z = 59.26$, $p < .001$).

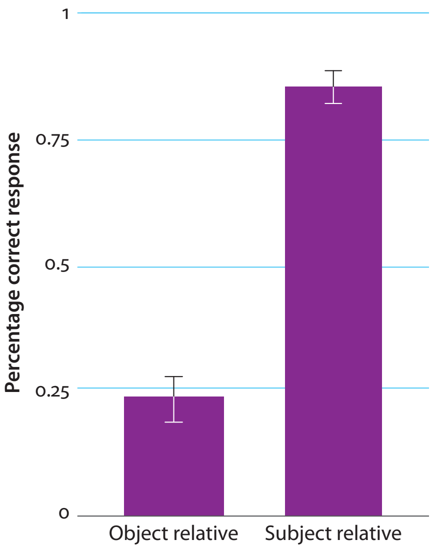

Figure 4. Accuracy proportion in Experiment 2

3.1.5 *Discussion*

Experiment 2 showed that participants were both slower at the spillover region and less accurate in answering the comprehension question in the object relative condition as compared to the subject relative condition. Importantly, comprehension accuracy results attested that participants failed to access the object reading 80% of the time, thus providing strong evidence that the late gender agreement cue was largely ineffective in triggering the object reanalysis. This result stands in stark contrast with results from Experiment 1, in which early cues failed to trigger reanalysis only 24% of the time. Moreover, since the percentage of correct responses in the OR condition in Experiment 2 did not differ from the percentage of incorrect responses in the SR condition, there is no evidence that participants ever truly accessed the object analysis when a late cue was provided.

Taken together, results from Experiments 1 and 2 suggest that late cues are poorly effective, if at all, in triggering reanalysis, unlike early cues. Given the striking asymmetry between the effectiveness of early and late cues in triggering reanalysis, we take these results to provide new evidence for *digging-in effects*. Specifically, the longer the parser has been committed with the incorrect analysis, the harder it is to reanalyze the sentence.

These effects are naturally captured by SOSP, which predicts later cues to be less effective than early cues in triggering reanalysis: linkages become stronger, and thus harder to undo, the longer the parser has been committed with the analysis. Our results are challenging for the cue-based memory model, which grants no role to the moment at which cues become available in predicting processing difficulties.

4. Conclusion

Italian relative clauses are ambiguous between a subject and object reading when the subject and the object display the same number. However, morpho-syntactic and syntactic cues such as word order and agreement favor disambiguation. The first aim of the present work was to collect direct evidence of the effectiveness of these cues in disambiguating the sentence, since previous studies in the literature based their conclusions either on reading time measures or on coarse-grained measures for adults, both of which provide little or no information about the parse that has been built. In Experiment 1, we collected a direct measure for reanalysis, i.e. comprehension accuracy. Results shown that comprehension accuracy was well above the chance level across conditions, indicating that the object analysis was successfully triggered most of the time, thus confirming the effectiveness of word order and agreement cues in triggering reanalysis in Italian ambiguous relative clauses.

The second aim of our study was to investigate whether the timing at which the disambiguating cue appears can influence reanalysis. In Experiment 2, we found that a late gender agreement cue presented after the RC was ineffective in triggering object reanalysis, unlike early number and word order cues in Experiment 1, which successfully triggered the object analysis well above the chance level. This result points to the key role played by the timing at which cues are made available in the input as a factor that determines the ease with which the comprehension system can revise an earlier commitment, in line with predictions from the SOSP model.

References

Adani, F., Van der Lely, H. K., Forgiarini, M., & Guasti, M. T. (2010). Grammatical feature dissimilarities make relative clauses easier: A comprehension study with Italian children. *Lingua*, 120(9), 2148–2166. https://doi.org/10.1016/j.lingua.2010.03.018

Arosio, F., Adani, F., & Guasti, M. T. (2009). Grammatical features in the comprehension of Italian relative clases by children. In J. M. Brucart, A. Gavarró, & J. Solà (Eds.), *Merging features: Computation, interpretation, and acquisition* (pp. 138–155). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199553266.003.0008

Bates, D., Martin M., Ben B., & Steve W. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Belletti, A., & Contemori, C. (2010). Disentangling the mastery of object relatives in children and adults. Evidence from Italian. In V. Moscati & E. Servidio (Eds.), *STiL - Studies in Linguistics* (CISCL Working Papers on Language and Cognition 4) (pp. 25–47). Siena: University of Sienna.

Belletti, A., Friedmann, N., Brunato, D., & Rizzi, L. (2012). Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua*, 122(10), 1053–1069. https://doi.org/10.1016/j.lingua.2012.02.007

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and language*, 3(4), 572–582. https://doi.org/10.1016/0093-934X(76)90048-1

De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-3184-1

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30, 725–745. https://doi.org/10.1016/0749-596X(91)90034-H

Frauenfelder, U., Segui, J., & Mehler, J. (1980). Monitoring around the relative clause. *Journal of Verbal Learning and Verbal Behavior*, 19(3), 328–337. https://doi.org/10.1016/S0022-5371(80)90257-1

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies* (Unpublished doctoral dissertation). University of Connecticut.

Frazier, L. (1987). Sentence processing. In M. Coltheart (Ed.), *Attention and Performance, XII*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, 119(1), 67–88. https://doi.org/10.1016/j.lingua.2008.09.002

Garraffa, M., & Grillo, N. (2008). Canonicity effects as grammatical phenomena. *Journal of Neurolinguistics*, 21(2), 177–197. https://doi.org/10.1016/j.jneuroling.2007.09.001

Grillo, N. (2009). Generalized minimality: Feature impoverishment and comprehension deficits in agrammatism. *Lingua*, 119(10), 1426–1443. https://doi.org/10.1016/j.lingua.2008.04.003

Guasti, M. T., Vernice, M., & Franck, J. (2018). Continuity in the adult and children's comprehension of subject and object relative clauses in French and Italian. *Languages*, 3(3), 24. https://doi.org/10.3390/languages3030024

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111(2), 228. https://doi.org/10.1037/0096-3445.111.2.228

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and causal- level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7(3), 376–395. https://doi.org/10.1162/jocn.1995.7.3.376

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: Tests in linear mixed effects models*. R package version 2.0-32.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, 47(1), 50–68. https://doi.org/10.1006/jmla.2001.2837

Mak, W. M., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language*, 54(4), 466–490. https://doi.org/10.1016/j.jml.2006.01.001

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, 29(2), 111–123. https://doi.org/10.1023/A:1005184709695

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48(1), 67–91. https://doi.org/10.1016/S0749-596X(02)00515-6

R Development Core Team. (2016). *R: A language and environment for statistical computing* [Computer Software]. Vienna. Austria: R Foundation for Statistical Computing.

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-prime user's guide. Pittsburgh*, PA: Psychology Software Tools.

Schriefers, H., Friederici, A. D., & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, 8, 499–520. https://doi.org/10.1006/jmla.1995.1023

Smith, G., & Tabor, W. (2018). Toward a theory of timing effects in self organized sentence processing. In I. Juvina, J. Houpt, & C. Meyers (Eds.), *Proceedings of ICCM 2018, 16th International Conference on Cognitive Modeling, 21-24 July, University of Wisconsin, Madison* (pp. 138–143). Madison, WI: University of Wisconsin.

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 431.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55(2), 157–166. https://doi.org/10.1016/j.jml.2006.03.007

Villata, S., Sprouse, J., & Tabor, W. (2019). Modeling ungrammaticality: A self-organizing model of islands. In A. K. Goel, C. M. Seifert, & C. Freska (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 1178–1184). Montreal, QB: Cognitive Science Society.

Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology*, 9, 2.

Empty categories

Brain responses elicited by implausible fillers and filled object gaps in German

Claudia Felser and Anna Jessen
University of Potsdam

We report the results from two experiments on the processing of filler-gap dependencies in German using event-related potentials (ERPs). Our aim was to identify and isolate brain responses linked to semantic vs. syntactic integration processes. Using maximally parallel stimulus materials, we examined ERPs at the subcategorising verb in sentences containing either a semantically implausible direct object filler or a filled direct object gap. Whilst the processing difficulty associated with trying to integrate an implausible filler with its subcategoriser was reflected in an enhanced N400 response, encountering a filled preverbal object gap elicited a P600 response at the verb but no negativity. These results confirm that the semantic and syntactic integration of displaced constituents are dissociable and qualitatively distinct processes.

Keywords: ERPs, filler-gap dependency, plausibility, filled gaps, German

1. Introduction

Sentences containing non-canonical word orders have been much discussed in the theoretical linguistic literature, and evidence from real-time processing studies has sometimes been put forward in support of specific theoretical hypotheses (e.g. Clahsen & Featherston, 1999; Gibson & Warren, 2004; Nakano, Felser, & Clahsen, 2002; Van de Koot, Silva, Felser, & Sato, 2015). In the sentence processing literature, sentences containing syntactically displaced constituents are usually referred to as filler-gap dependencies. Encountering a fronted *wh*-phrase or 'filler' such as *which student* in (1) will trigger a search for a corresponding 'gap', that is, for some point in the sentence at which the filler can be properly integrated.[1]

1. Note that, although for expository reasons we have indicated the filler's canonical structural position using underscores, the term 'gap' is intended to be theory-neutral in that it does not

https://doi.org/10.1075/lald.64.04fel

(1) Which student did you meet __ at the party?

Coming across a potential lexical licenser such as the verb *meet*, which is transitive but lacks a post-verbal direct object, will result in the processor establishing a link between the filler and the verb such that the filler is interpreted as the verb's direct object and theme argument. The processor's tendency to try and link a filler to the first grammatically possible gap is known as the *Active Filler Strategy* (Clifton & Frazier, 1989).

Exactly how such a link is established is still controversial, however. One possibility is that a semantic link is established between a filler and a potential lexical licenser (e.g. Pickering & Barry, 1991). Under this view, upon encountering a verb such as *meet* in (1), which normally requires a theme argument, a filler that is a potential theme (such as *which student*) will be integrated directly into the verb's thematic grid. Alternatively, filler integration may involve an additional syntactic step, with the filler being linked to a vacant phrase-structure slot corresponding to its canonical structural position (e.g. Gibson & Hickok, 1993).

In most previous studies on the processing of filler-gap dependencies, these two possibilities were difficult to dissociate empirically, but there is some evidence from the processing of indirect object dependencies suggesting that both lexical-semantic and syntactic integration processes are involved (Jessen, Festman, Boxell, & Felser, 2017; Nicol, 1993). In these studies, effects of filler integration difficulty were observed both at the filler's lexical-semantic licenser (a ditransitive verb) and at a non-adjacent indirect object gap further downstream. Nicol (1993), for example, reported evidence from cross-modal priming showing that a fronted prepositional phrase (such as *to which doctor* in 2) was mentally reactivated both at the subcategorising verb (*give*) and at the filler's canonical structural position following the presentation of the direct object (*a long thick winter coat*).

(2) To which doctor did the bricklayer from a small community in North Dakota give a long thick winter coat __ a few weeks ago?

In languages such as English in which verb phrases are head-initial (VO languages), lexically driven integration effects will necessarily precede any syntactic gap-filling effects; it remains unclear to what extent the integration processes involved are indeed independent, however.

Filler-gap processing studies on languages with head-final verb phrases (OV languages) have shown that putative object gaps may be postulated even before the subcategorising verb has been fully processed (e.g. Clahsen & Featherston, 1999; Fiebach, Schlesewsky, & Friederici, 2002; Nakano et al., 2002), but none

necessarily refer to a position in a phrase marker.

of these studies were designed to isolate and compare different filler integration sub-processes. In the current study we seek to empirically dissociate semantic and syntactic integration processes by investigating readers' brain responses both to implausible fillers and to filled direct object gaps in German.

2. Theoretical background

Recording readers' or listeners' brain responses to linguistic stimuli (event-related potentials, or ERPs) allows us to tap into their immediate and largely unconscious reactions to particular properties of these stimuli (see Rommers & Federmeier, 2017, for an overview). During the past few decades, several ERP components have been identified that covary with linguistic properties. One of these is the N400, a negative-going wave form which peaks around 400 ms after stimulus onset. The N400 is a normal brain response to meaningful words whose amplitude is inversely linked to the degree to which a word fits the preceding context; hence an enhanced N400 is often described as indexing semantic anomaly. Another well-known ERP response is the P600, a positive-going wave form that is typically elicited by syntactic anomaly but which has also been described as indexing syntactic processing difficulty more generally. A third kind of brain response are left-anterior negativities (LANs), which have been linked to morphosyntactic processing but also to working memory processes.

Several previous ERP studies examining the processing of filler-gap dependencies have sought to identify potentially distinct sub-processes involved in filler-gap dependency resolution. The results from some studies indicate that memory storage and filler integration are reflected in different ERP components, a LAN and a P600, respectively (Felser, Clahsen, & Münte, 2003; Fiebach et al., 2002; Phillips, Kazanina, & Abada, 2005). Here we investigate whether the integration process itself can be broken down into qualitatively distinct subprocesses.

Manipulating a filler's semantic fit with a potential lexical licenser can serve as an experimental diagnostic for lexically-based filler integration. Garnsey, Tanenhaus, and Chapman (1989), for example, compared English speakers' ERPs to sentences such as (3), which contained either a plausible (*which customer*) or implausible (*which article*) direct object filler.

(3) The business man knew which customer/#article the secretary called __ at home.

For sentences containing an implausible filler, encountering the embedded verb (*called*) elicited an N400 response. This shows that trying to integrate a potential object filler that is not a plausible theme argument of the verb gives rise to greater processing difficulty compared to a filler that is a good semantic fit (i.e. with higher predictability).

Syntactic integration difficulty, on the other hand, can be revealed by so-called 'filled-gap' effects. These are observed when a syntactic position that might host a filler is found to be already occupied (Stowe, 1986). Whilst semantically or pragmatically anomalous fillers typically elicit an N400 response (see also Dallas, DeDe, & Nicol, 2013), the picture is less clear for filled-gap violations. Encountering a filled gap may either be perceived as a phrase-structure or as a semantic violation, with a 'superfluous' constituent competing with the filler for the same thematic role.

Friederici and Frisch (2000) presented readers with ungrammatical sentences that contained an intransitive verb followed by a superfluous argument phrase such as *den Geiger* 'the violonist' in (4).

(4) * *Heute trödelte der Cousin den Geiger am Aufzug*
today dawdled the cousin$_{\text{NOM}}$ the violinist$_{\text{ACC}}$ at.the elevator

Encountering the extra argument phrase elicited a biphasic ERP response, an N400 followed by a P600. The authors interpreted the N400 as reflecting thematic integration difficulty and the subsequent P600 as reflecting the parser's attempt to repair the sentence's syntactic representation, following its failure to integrate the extra argument phrase into the verb's thematic grid. Note, however, that the authors did not measure ERPs at the verb itself.

Hestvik, Maxfield, Schwartz and Shafer (2007) investigated filled-gap effects in English sentences containing object relative clauses (RCs). Their auditorily presented stimulus materials included ungrammatical sentences such as (5a) below. Here the modified matrix subject *the zebra* is likely to be construed as the direct object of the verb *kissed* when the verb is first encountered. Finding the post-verbal position to be already filled by another noun phrase (*the camel*) is expected to yield filled-gap effects.

(5) a. * The zebra that the hippo kissed <u>the camel</u> on the nose ran far away.
b. The zebra said that the hippo kissed <u>the camel</u> on the nose and then ran far away.

Coming across a potential gap position that was filled by another noun phrase (such as *the camel* in 5a) gave rise to a biphasic brain response, an early left-anterior negativity (ELAN) and a later P600, with the latter however not reaching statistical significance. Given that an ELAN is sometimes observed in response to phrase-structure violations (Friederici, 2002), the authors interpreted this as evidence that filled gaps are perceived as phrase-structure violations, with two constituents competing for the same syntactic slot. In a follow-up study, Hestvik, Bradley, and

Bradley (2012) found that filled object gaps triggered a bilateral anterior negativity and a subsequent P600. Both ERP effects were modulated by participants' scores in a working memory (WM) test, with high-WM participants showing a significant anterior negativity in an earlier time window than the low-WM participants, as well as an earlier significant P600 component. Neither study reported any N400 effects, and the functional significance of the negative-going ERPs observed by Hestvik and colleagues is not entirely clear (see Steinhauer & Drury, 2012, for a critical discussion of the ELAN component).

In an attempt to dissociate potential verb-driven effects from position-specific gap-filling effects, Jessen et al. (2017) examined the processing of indirect object dependencies in English, also using a filled-gap paradigm. As noted above, using indirect object fillers allows for semantic integration effects at a ditransitive verb to be empirically distinguishable from syntactic integration effects at the filler's canonical structural position, as this is non-adjacent to the verb (Nicol, 1993).

Participants' brain responses were recorded while they read ungrammatical sentences such as (6a) below, which contained an indirect object filler (such as *for which*, referring to *the monkey*) but a filled indirect object gap in the RC, in comparison to non-filler-gap control sentences like (6b).

(6) a. *Sarah tickled the monkey for which Peter arranged some classes <u>for it</u> after the vacation.
b. Sarah tickled the monkey while Peter arranged some classes <u>for it</u> after the vacation.

Jessen et al. (2017) found that sentences containing a *wh*-filler such as (6a) elicited an enhanced N400 response to a ditransitive verb (*arranged*), indicating that participants tried to link the filler semantically to the first potential lexical licenser they came across. At the filler's canonical structural position, finding this position filled by another constituent (such as the prepositional phrase *for it* in 6a) elicited a P600 response. These findings support the claim that filler integration is a two-step process (Nicol, 1993), with semantic and syntactic integration effects being isolable in the same sentence. However, given that the N400 and P600 effects in Jessen et al.'s (2017) study were elicited by different words, more research is needed to corroborate this conclusion.

3. The current study

We carried out two complementary experiments investigating the processing of direct object filler-gap dependencies in German. Note that German is considered to be an OV language despite the fact that German main clauses require finite verbs to appear in the second structural position (e.g. Bach, 1962), which is why direct object gaps are thought to precede their subcategorising verb. The OV pattern is preserved in embedded clauses. Whilst in previous studies on English, effects of implausible fillers were measured at the verb, filled gap effects could only be measured at a postverbal noun phrase or prepositional phrase; however, the OV property of German makes it possible to measure both plausibility and filled-gap effects at the same point, the subcategorising verb.

Testing the same group of participants on sentences containing either an implausible object filler or a filled direct-object gap allows us to neutralize possible individual factors such as WM limitations that may affect filler-gap processing (compare e.g. Hestvik et al., 2012; Nakano et al., 2002). To further maximize between-experiment comparability, we kept the syntactic structure of our stimulus sentences as similar as possible in both experiments.

Based on previous findings our predictions were as follows:

1. Sentences containing a semantically implausible object filler should elicit an N400 response at the filler's potential lexical-semantic licenser.
2. Sentences containing a filled object gap should elicit a P600 once the filled gap has been identified. If encountering a filled gap also triggers semantic integration problems at the verb (i.e., thematic role competition), we might find the P600 preceded by an N400. If, on the other hand, a filled gap constitutes a phrase-structure violation, we might observe an additional (left-)anterior negativity.

Materials

Plausibility manipulation

We created 48 sentence pairs as in Example (7). Our experimental stimuli all contained a main clause headed by a transitive verb, followed by an object relative clause that was introduced by a case and gender-unambiguous relative pronoun. The RC always modified the noun phrase (NP) immediately preceding it, which functioned as the direct object of the verb in the main clause. Both the relative pronoun and its antecedent NP (henceforth referred to as the 'filler') were marked for accusative case. An accusative-marked relative pronoun indicates the presence of a direct object gap in the RC. The filler NP was either a plausible (7a) or an implausible (7b) direct object of the RC.

(7) a. *Kristin bekam den Brief, den der Geliebte __ gelesen hatte ohne*
Kristin received the letter which the lover read had without
jede Erlaubnis.
any permission

b. # *Kristin bekam den Kater, den der Geliebte __ gelesen hatte ohne*
Kristin received the tomcat which the lover read had without
jede Erlaubnis.
any permission
'Kristin received the letter/tomcat that the lover had read without permission.'

The RC always contained a transitive verb in its past participle form, followed by the auxiliary *hatte* 'had' and an adverbial adjunct which served as an additional "padding" phrase between the critical verb and the end of the sentence.

The object nouns of the main clause (e.g. *Brief/Kater*) were carefully matched for length (number of letters) and frequency (according to the CELEX data base; Baayen, Piepenbrock, & Van Rijn, 1993). The mean length of our object nouns was 5.92 letters (SD = 1.9) in the plausible and 5.98 letters (SD = 1.96) in the implausible condition. A paired sample t-test showed no significant between-condition difference in word length ($t = -0.15$, df = 94, $p = .8$). The mean frequency of plausible object nouns was 25.7 per million (SD = 43.4) and 40.1 (SD = 50.5) for implausible ones. This difference was also non-significant ($t = -1.5$, df = 94, $p = .1$). On order to control for possible animacy effects (Mak, Vonk, & Schriefers, 2002), half of the plausible fillers were inanimate and half were animate nouns, and vice versa for the implausible fillers.

Filled-gap manipulation

A further set of 48 grammatical/ungrammatical sentence pairs were created as shown in Example (8).

(8) a. *Philipp nahm den Knochen, obwohl die Halterin ihrem Hund <u>den</u>*
Philipp took the bone although the owner her dog it
gegeben hatte.
given had
'Philipp took the bone although the owner had given it to her dog.'

b. * *Philipp nahm den Knochen, den die Halterin ihrem Hund <u>den</u> gegeben*
Philipp took the bone which the owner her dog it given
hatte.
had
'Philipp took the bone which the owner had given to her dog.'

Grammatically well-formed sentences such as (8a) that did not contain a filler-gap dependency served as control items for our filled-gap sentences (8b). All experimental sentences consisted of a main clause followed by a subordinate clause and contained a pronominal determiner (henceforth, d-pronoun) at the embedded verb's canonical direct object position.

Our control items (8a) contained an adverbial clause introduced by a subordinating conjunction including *weil* ('because', $n = 15$), *nachdem* ('after', $n = 14$), *obwohl* ('although', $n = 10$) and *bevor* ('before', $n = 9$). As noted by a reviewer, the position of the d-pronoun *den* in (8a) is somewhat marked, because weak or unstressed pronouns normally precede the indirect object (Müller, 2001). What matters here, however, is that (8a) contrasts clearly with the ungrammaticality of (8b), which contains a filled gap.

The filled-gap sentences (8b) differed from their well-formed counterparts in that they contained an object RC modifying the object NP in the main clause. They were thus structurally similar to the plausible/implausible sentence pairs described above. Our filled-gap sentences were grammatically ill-formed, however, because they contained a superfluous d-pronoun at the position of the purported gap. We used d-pronouns rather than full noun phrases because the former are semantically underspecified, so that ERP effects reflecting the parser's attempt to integrate the d-pronoun into the verb phrase should not be affected by their lexical-semantic or pragmatic compatibility. Although the embedded clauses in sentences like (8b) do not provide an environment that licenses resumptive pronouns, another reason for using d-pronouns was that rather than giving rise to argument competition (compare e.g. Hestvik et al., 2007) they could possibly be interpreted resumptively. That is, using d-pronouns in our filled-gap condition rendered these sentences ungrammatical but not necessarily uninterpretable.

Note that in (8b), the d-pronoun *den* could potentially introduce an adverbial expression (such as *den Tag zuvor* 'the day before'). In our experimental sentences, this ambiguity is resolved at the verb following it (e.g. the participle *gegeben* 'given'), at which point the ill-formedness should be obvious and which was also our point of measurement. In 23 (out of 48) sentence pairs the subordinate clause subject and following possessive pronoun were feminine, the rest were masculine.

Our stimulus items were distributed across two lists so that each list contained 24 plausible and 24 implausible items, and 24 filled-gap (ungrammatical) and 24 control items without a gap (grammatical), but never both members of the same minimal sentence pair. Our 96 experimental sentences were mixed with 60 distractor items with varying grammatical and semantic structures. Each presentation list thus included a total of 156 sentences. One third of all the sentences ($n = 52$) were followed by a yes/no comprehension question (e.g. "The dog gave the owner the bone?" for Example 8 above) to help ensure that participants read the sentences carefully for comprehension.

Participants

We tested 21 German native speakers (10 female, mean age 23.2 years, range 18–33). Participants were recruited from Potsdam University and the surrounding area. None of them reported to have grown up bilingually, and they had all started learning other languages at school. All were right handed, had normal or corrected to normal vision, and had no history of developmental and/or language disorders.

Participants were tested on both presentation lists but during separate testing sessions. The mean time-window between testing sessions was 11.6 days (SD 10.2).

Procedure

The experiment took place in a quiet laboratory specifically designed for EEG experiments. Prior to the experiment participants filled in a biographical questionnaire including information about their language acquisition history, received information about the EEG procedure and signed a consent form. Participants received either a small fee or course credit for their participation.

The EEG experiment was programmed using Presentation version 14.9 (Neurobehavioral Systems). Sentences were presented visually word-by-word with 600 ms inter-stimulus intervals. Words were presented on a 61cm computer monitor in comic sans 96-point size font in black against white background. End-of-trial questions were presented in green letters (40 points size font) against white background and remained on the screen until participants answered them via a button press. Participants received feedback in the form of either a smiley face for correct answers, or a blank screen if an answer was incorrect, which was presented for 1000 ms. The appearance of the end-of-trial questions was pseudo-randomized over items (controlled for amount of questions per condition) and participants.

Each trial began with a fixation cross presented for 100 ms and ended with a 500 ms black screen. Before the start of the experiment, participants were given five practice trials, each followed by a question. If they had no further questions about the task, the experiment started. Each session lasted about 80 minutes including electrode set up and hair washing afterwards. The experiment itself had a duration of about 30 minutes and contained several breaks, which participants could end at their will.

The experiments were approved by the University of Potsdam ethics committee (application number 37/2011).

EEG recording and analysis

Brain waves were recorded continuously from 31 electrodes plus an additional vertical electro-occulogram underneath the right eye to detect blinks (Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CP2, CP6, TP10, P7, P3, Pz, P4, P8, PO9, O1, O2, PO10), using BrainVision Recorder software. Signals were recorded with an on-line band-pass filter between 0.016 70Hz and digitized at 500Hz. Electrode impedances were kept below 20kΩ (in accordance with the guidelines for using ActiCaps). The reference electrode was FCz, but electrodes were offline re-referenced to the average of the left and right mastoids. Recordings were offline filtered between 0.1 Hz and 30 Hz. The EEG data were processed with BrainVision Analyzer 2. For single-segment analysis, epochs were extracted from 200 ms before the presentation of a critical word (the participle in the subordinate clause) until 1200 ms after the presentation onset, resulting in 1400 ms epochs (–200 ms to 1200 ms), and were baseline-corrected using a 200 ms pre-stimulus baseline, following the recommendations by Tanner, Norton, Morgan-Short, & Luck (2016).

To remove typical muscle and eye movement artifacts from the recordings, an independent component analysis (ICA) algorithm (Infomax) was applied to the data. Epochs containing additional artifacts were identified with the semi-automatic rejection option in the Analyzer and rejected after visual inspection. Both data sets from one participant as well as three further (single) data sets were excluded from analysis due to overall noisy data. After artifact rejection, 98.1% of all trials from the remaining participants (37 data sets in total from 20 participants) were included in the statistical analysis.

For statistical analysis, 21 electrodes were pooled into five regions of interest: Fz. Cz, Pz (midline), F7, F3, FC5, FC1, C3 (left frontal), F8, F4, FC6, FC2, C4 (right frontal), CP5, CP1, P7, P3 (left posterior), CP6, CP2, P8, P4 (right posterior). This is a conservative and fairly common way of defining ROIs in ERP studies (compare e.g. Jessen & Felser, 2019). Time windows of interest were selected based on a 50 ms time line analysis and visual inspection.

Our analysis included the factors region of interest (ROI), condition (plausible vs. implausible in the first experiment, and grammatical vs. ungrammatical in the second) as within-participants factors, and testing session (1st vs. 2nd) as a between-participants factor, since from three of our participants we could only use a single data set. Further stepping-down analyses by ROI contained condition as a within-participants and testing session as a between-participants factor. All statistical analyses were carried out using R version 3.4.1 and the ezANOVA function of the ez package.

Results

Behavioural results. Participants responded at 80.9% (range 69.8–93.8%) correct to the yes/no comprehension questions. This indicates that they read the stimulus material actively for comprehension.

ERP results for implausible fillers. Participants' brain responses to the critical verb show a globally distributed negativity for implausible over plausible sentences, as can be seen in Figure 1. In a time-window from 400–600 ms after verb onset (e.g. *gelesen* in Example 7) there were significant main effects of ROI [$F(4,140) = 13.8$, $p < .001$], and condition [$F(1,35) = 4.38$, $p = .04$] but no interaction between the two [$F(4,140) = 0.08$, $p = .9$]. We found no main effect for testing session [$F(1,35) = 0.44$, $p = .5$] and no interaction between testing session and condition [$F(1,35) = 0.35$, $p = .55$].

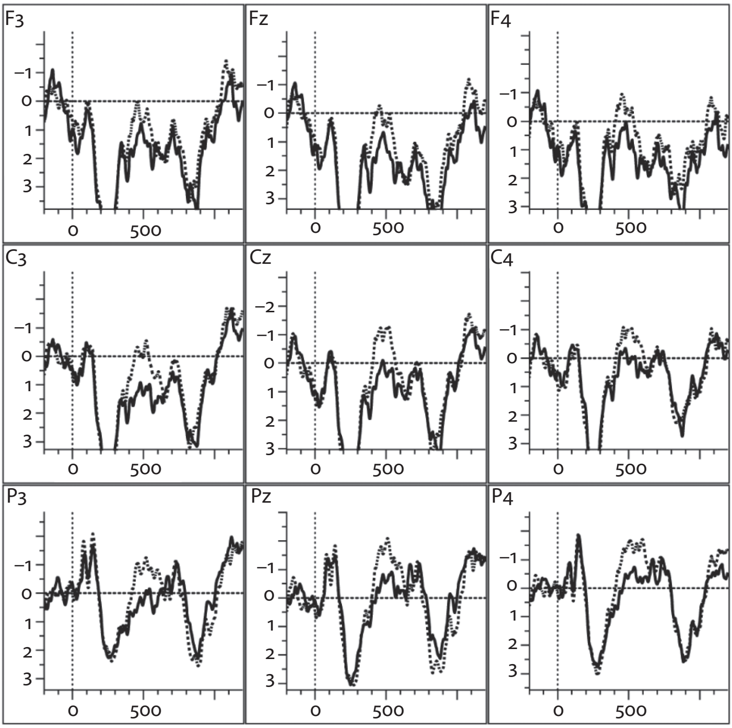

Figure 1. Grand average ERPs at the verb for sentences containing implausible (dotted line) vs. plausible (solid line) fillers. Negativity is plotted upwards. The vertical dotted line indicates the verb's onset.

A closer look into individual ROIs revealed significant between-condition differences in the midline [$F(1,35) = 4.68$, $p = .04$] as well as the right posterior ROI [$F(1,35) = 4.47$, $p = .04$], and marginal effects in the right frontal [$F(1,35) = 3.4$, $p = .07$] and the left posterior ROIs [$F(1,35) = 3.62$, $p = .065$]. In none of the ROIs did the factor testing session interact with condition (midline: $F(1,35) = 0.2$, $p = .7$; right posterior: $F(1,35) = 0.05$, $p = .8$; right frontal: $F(1,35) = 0.39$, $p = .5$; left posterior: $F(1,35) = 0.45$, $p = .5$). These results confirm that the effect is globally distributed but most pronounced over the right posterior region, and therefore shows the typical distribution and timing of an N400.

ERP results for filled gaps. We again chose the embedded participle (*gegeben* 'given' in Example 8) as the critical analysis region. This was because of the temporary ambiguity of the preceding d-pronoun *den*, and to ensure maximum comparability between our two experiments. In a time-window from 400–700 ms after verb onset a positivity for the ungrammatical filled-gap sentences emerged over posterior electrodes (see Figure 2). Statistical analysis revealed a main effect for ROI [$F(4,140) = 3.35$, $p = .01$] as well as an interaction of ROI with condition [$F(4,140) = 2.7$, $p = .03$]. The factor first or second test session showed a marginal main effect [$F(1,35) = 3.6$, $p = .07$], but did not interact with condition [$F(1,35) = 0.48$, $p = .5$], nor with ROI and condition [$F(4,140) = 0.86$, $p = .5$].

Stepping down by ROI revealed a significantly more positive-going wave form for the ungrammatical sentences in the right posterior [$F(1,35) = 6.79$, $p = .01$] and the left posterior ROIs [$F(1,35) = 8.2$, $p = .007$]. The factor testing session did

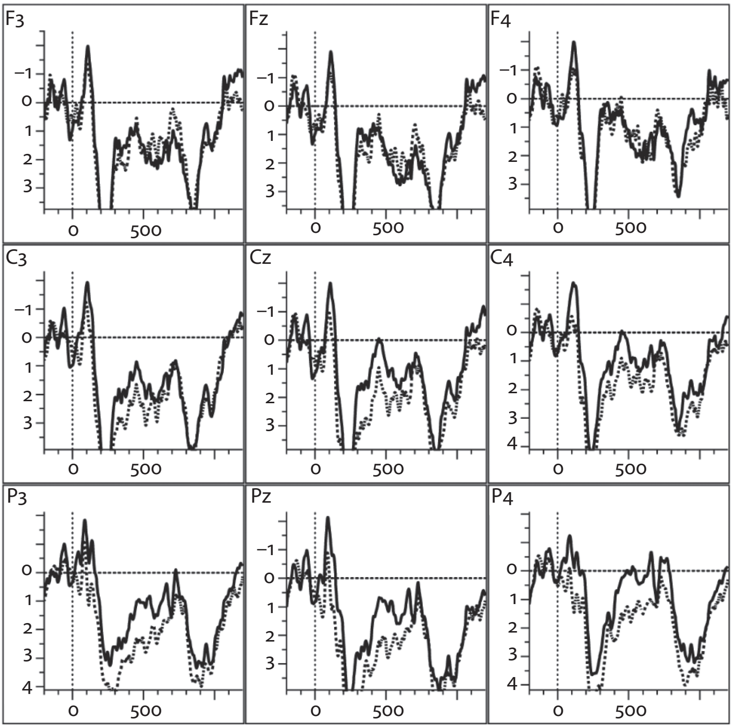

Figure 2. Grand average ERPs at the verb for sentences containing a filled gap (dotted line) vs. no gap (solid line). Negativity is plotted upwards. The vertical dotted line indicates the verb's onset

not interact with condition in either ROI (right posterior: $F(1,35) = 0.02$, $p = .88$; left posterior: $F(1,35) = 0.7$, $p = .4$). No other region of interest showed any significant effects. The timing and distribution of the observed positivity is typical of the P600 component.

We additionally looked at the actual filled gap (the d-pronoun *den* preceding the participle), but did not find any between-condition differences for this word at all, as expected.

4. Discussion

Using structurally similar stimulus sentences, our two experiments revealed distinct ERP responses elicited by implausible fillers vs. filled gaps. At the subcategorising verb, implausible fillers elicited an N400 response whilst a direct object gap filled by a d-pronoun elicited a P600 response. Let us first consider the findings from our plausibility manipulation. Given that sentences such as (7b) (repeated in 9 below for convenience) are semantically odd but syntactically well-formed, we follow Garnsey et al. (1989) and others in interpreting the N400 as an index of semantic integration difficulty.

(9) # *Kristin bekam den Kater, den der Geliebte __ gelesen hatte ohne jede*
Kristin received the tomcat which the lover read had without any
Erlaubnis.
permission
'Kristin received the tomcat that the lover had read without permission.'

This finding provides further confirmation of the Active Filler Strategy (Clifton & Frazier, 1989) and is line with the results from ERP studies on English filler-gap dependencies which also found an enhanced negativity for implausible object fillers (e.g. Jessen & Felser, 2019; Dallas et al., 2013).

In contrast to semantically anomalous sentences such as (9) above, for filled-gap sentences of the kind we used in the current study, the addition of a superfluous d-pronoun rendered our sentences ungrammatical but not necessarily uninterpretable. Even though P600 responses have sometimes been reported to index semantic anomaly (e.g. Bornkessel-Schlesewsky, & Schlesewsky, 2008; Van Herten, Kolk, Chwilla, 2005), they have traditionally – and very robustly – been linked to syntactic processing difficulty (Osterhout & Holcomb, 1992; Hagoort, Brown, & Groothusen, 1993; Kaan, Harris, Gibson, & Holcomb, 2000, among many others). Although the number of published studies that have examined listeners' or readers' brain responses to filled gaps is still relatively small, they consistently report a P600 (but no N400) in response to a filled gap (Hestvik et al., 2007; Hestvik et al., 2012; Jessen et al., 2017). The results from our filled-gap experiment are in line with these findings and provide compelling evidence that P600 effects elicited by filled gaps do indeed reflect syntactic integration difficulty.

The fact that in our filled-gap sentences the object position was filled by a d-pronoun rather than by a fully specified noun phrase makes it very unlikely that the P600 effect we observed reflects semantic integration problems. If the d-pronoun were interpreted as a resumptive pronoun referring to the filler *den Knochen* ('the bone'), then its presence would not in fact give rise to any thematic-role competition. The 'resumptive' use of the d-pronoun *den* is illustrated in (10), where it occupies a position at which a gap would not be licensed. Here the gap associated with the fronted prepositional phrase *von dem* 'of which' is adjacent to the RC's main verb *wusste* 'knew', whilst the object of the more deeply embedded verb *gegeben* 'given' is expressed as a d-pronoun (also referring to *den Knochen* 'the bone').

(10) *Philipp nahm den Knochen, [$_{PP}$ von dem] er __ wusste, dass die Halterin* <u>*den*</u>
Philipp took the bone of which he knew that the keeper it
ihrem Hund gegeben hatte.
her dog given had
'Philipp took the bone which he knew that the keeper had given to her dog.'

In this respect our stimulus materials resembled those used by Jessen et al. (2017) (compare example 6 above) but differed from those used by Hestvik et al. (2007) and Hestvik et al. (2012), whose filled-gap stimuli contained full NPs in direct object position that were semantically distinct from the filler and thus in clear competition with it. Our use of a semantically underspecified d-pronoun that could potentially be interpreted resumptively, instead of a full lexical NP, might be one reason why we observed a P600 but no additional frontal negativity as was observed by Hestvik and colleagues.[2]

Taken together, the present findings confirm that the semantic and syntactic integration processes involved in establishing filler-gap dependencies are separable and independent.

Acknowledgements

This research was supported by an Alexander-von-Humboldt professorship awarded to Harald Clahsen. We thank Caroline Beese for help with the materials creation and programming, Julia Festman for helpful advice, and Michela Mosca for help with the data collection.

2. Another difference between ours and previous studies is our use of visual rather than auditory stimuli: ELANs are more likely to be elicited by auditory stimuli (Steinhauer & Drury, 2012).

References

Baayen, H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX Lexical Database (CD ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bach, E. (1962). The order of elements in a transformational grammar of German. *Language*, 38, 263–269. https://doi.org/10.2307/410785

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, 59, 55–73. https://doi.org/10.1016/j.brainresrev.2008.05.003

Clahsen H., & Featherston S. (1999) Antecedent priming at trace positions: Evidence from German scrambling. *Journal of Psycholinguistic Research*, 28, 415–437. https://doi.org/10.1023/A:1023293132656

Clifton, C. Jr., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In M. Tanenhaus & G. Carlson (Eds.), *Linguistic structure in language processing* (pp. 273–317). Dordrecht: Springer. https://doi.org/10.1007/978-94-009-2729-2_8

Dallas, A., DeDe, G., & Nicol, J. (2013). An event-related potential (ERP) investigation of filler-gap processing in native and second language speakers. *Language Learning*, 63, 766–799. https://doi.org/10.1111/lang.12026

Felser, C., Clahsen, H., & 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 and Language*, 87, 345–354. https://doi.org/10.1016/S0093-934X(03)00135-4

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, 47, 250–272. https://doi.org/10.1016/S0749-596X(02)00004-9

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6, 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Friederici, A. D., & Frisch, S. (2000). Verb argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and Language*, 43, 476–507. https://doi.org/10.1006/jmla.2000.2709

Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research*, 18, 51–60. https://doi.org/10.1007/BF01069046

Gibson E., & Hickok, G. (1993). Sentence processing with empty categories. *Language and Cognitive Processes*, 8, 147–161. https://doi.org/10.1080/01690969308406952

Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, 7, 55–78. https://doi.org/10.1111/j.1368-0005.2004.00065.x

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439–483. https://doi.org/10.1080/01690969308407585

Hestvik, A., Bradley, E., & Bradley, C. (2012). Working memory effects of gap-predictions in normal adults: An event-related potentials study. *Journal of Psycholinguistic Research*, 41, 425–438. https://doi.org/10.1007/s10936-011-9197-8

Hestvik, A., Maxfield, N., Schwartz, R. G., & Shafer, V. (2007). Brain responses to filled gaps. *Brain and Language*, 100, 301–316. https://doi.org/10.1016/j.bandl.2006.07.007

Jessen, A., & Felser, C. (2019). Reanalysing object gaps during non-native sentence processing: Evidence from ERPs. *Second Language Research*, 35, 285–300. https://doi.org/10.1177/0267658317753030

Jessen, A., Festman, J., Boxell, O., & Felser, C. (2017). Native and non-native speakers' brain responses to filled indirect object gaps. *Journal of Psycholinguistic Research*, 46, 1319–1338. https://doi.org/10.1007/s10936-017-9496-9

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15, 159–201. https://doi.org/10.1080/016909600386084

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, 47, 50–68. https://doi.org/10.1006/jmla.2001.2837

Müller, G. (2001). Order preservation, parallel movement, and the emergence of the unmarked. In G. Legendre, J. Grimshaw, & S. Vikner (Eds.), *Optimality-theoretic syntax* (pp. 279–313). Cambridge, MA: The MIT Press.

Nakano, Y., Felser, C., & Clahsen, H. (2002). Antecedent priming at trace positions in Japanese long-distance scrambling. *Journal of Psycholinguistic Research*, 31, 531–571. https://doi.org/10.1023/A:1021260920232

Nicol, J. (1993). Reconsidering reactivation. In G. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 321–350). Hillsdale, NJ: Lawrence Erlbaum Associates.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, 6, 229–259. https://doi.org/10.1080/01690969108406944

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22, 407–428. https://doi.org/10.1016/j.cogbrainres.2004.09.012

Rommers, J., & Federmeier, K. D. (2017). Electrophysiological methods. In A. M. B. de Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics and the neurobiology of language: A practical guide* (pp. 247–265). Hoboken, NJ: Wiley Blackwell.

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, 120, 135–162. https://doi.org/10.1016/j.bandl.2011.07.001

Stowe, L. A. (1986). Parsing *wh*-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1, 227–245. https://doi.org/10.1080/01690968608407062

Tanner, D., Norton, J. J., Morgan-Short, K., & Luck, S. J. (2016). On high-pass filter artifacts (they're real) and baseline correction (it's a good idea) in ERP/ERMF analysis. *Journal of Neuroscience Methods*, 265, 166–170. https://doi.org/10.1016/j.jneumeth.2016.01.002

Van de Koot, H., Silva, R., Felser, C., & Sato, M. (2015). Does Dutch A-scrambling involve movement? Evidence from antecedent priming. *The Linguistic Review*, 32, 739–776. https://doi.org/10.1515/tlr-2015-0010

Van Herten, M., Kolk, H., & Chwilla, D. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22, 241–255. https://doi.org/10.1016/j.cogbrainres.2004.09.002

Gone with a trace?

Reactivation at PRO positions

Tori Larsen and Christer Johansson
University of Bergen

PRO is a theoretically attractive empty category proposed to fill the subject position in non-finite clauses. This chapter is concerned with the empirical evidence for PRO. We investigate reactivation patterns in Control sentences in Norwegian with data from two reaction time experiments. We use a picture recognition task to measure the time and position of PRO antecedent reactivation during sentence processing. The data analysis explores a baseline that has the ability to separate syntactic priming (correct PRO antecedent reactivation) and repetition priming. We find a significant effect of position and an effect of syntactic priming for Subject Control using mixed effects models. We discuss possible theoretical implications and necessary future research.

Keywords: control, PRO, coreference, trace, reactivation, priming, Norwegian, sentence processing

1. Introduction

The existence of null items in linguistic theory is controversial. Although such items may improve the beauty and symmetry of syntactic theory, sometimes they create more issues than they resolve. Often, they lead to predictions that can be tested by observing their effects on human language processing. For syntactic traces that result from movement, there is robust research illustrating reactivation effects (Nicol, 1988; Nicol & Osterhout, 1988; Nicol & Swinney, 1989; Hestvik et al., 2005, 2010; inter alia). These effects show that processing is affected in predictable ways, such as by immediate reactivation of a trace referent at the predicted trace position. In these instances, even if there is no overt lexical element at the trace position, it is possible to observe measurable effects for recognition that are consistent with the notion of a salient, covert item occupying that position.

https://doi.org/10.1075/lald.64.05lar

PRO is different than an NP trace. This placeholder does not result from movement but is motivated by uniformity of theta role (functional role) assignment (Chomsky, 1981/1993). PRO can be *coreferent* with earlier sentence material, as in instances of Obligatory Control (cf. Rosenbaum, 1967; Radford, 2004). It can also be optionally controlled, or arbitrarily controlled by a noun phrase (NP) outside of the sentence. The Extended Projection Principle (EPP), presented by Chomsky (1982), assumes that all major verb phrase (VP) projections distribute a subject role. This goes for non-finite clauses as well, though they never contain an overt subject. Consequently, the possibility of an overt subject must be blocked, and so PRO fills that position for the non-finite verb. The EPP demands that all clauses have a subject; therefore, PRO becomes the covert subject in non-finite clauses. An example of a Control sentence taken from Carnie (2013) is shown in (1).

(1) Jean persuaded Robert$_{i}$ PRO$_{i}$ to leave.

However, PRO itself does not necessarily have an antecedent and has characteristics of both a pronoun and an anaphor. Due to these exceptional characteristics, the PRO Theorem excludes it from the Binding Theory, concluding that PRO must be ungoverned (Chomsky, 1981/1993). Chomsky and Lasnik (1993) also argue that PRO has Null Case under this theorem (in English) (cf. Martin, 2001). Sigurðsson (2008) shows that agreement in Icelandic indicates that PRO may agree with Cases other than Null Case. This is possibly because it agrees with the subject Case of the main verb in the non-finite clause. This further supports the existence of a covert element, PRO, but lends to the mystery of its syntactic properties.

There is one more specific and intriguing controversy that our experiment may shed some light on. It is thought that the non-finite tense marker licenses PRO. It is debated whether PRO undergoes movement to the spec-TP position, or if it stays at the spec-VP position where it originates, after the infinitive marker. Baltin (1995) argues that the tense marker *does not* trigger movement, while others, such as Radford (2004), claim that PRO *must* move to spec-TP. Yet another theory proposes that PRO holds a position somewhere in between spec-TP and spec-VP, possibly spec-AuxP (Reed, 2014). These theories of PRO position yield (at least) two different structures, illustrated below in examples (2) and (3). Another option is that PRO is activated in both positions, if PRO moves and leaves a trace.

(2) [$_{TP}$ [$_{T}$ to] [$_{VP}$ PRO]]

(3) [$_{TP}$ PRO$_{i}$ [$_{T}$ to] [$_{VP}$ t_{i}]]

Examples (2) and (3) show the position of PRO before and after the English infinitive marker *to*. Our research is on Norwegian, but the linear positioning of PRO is assumed to be the same in both languages. Though the syntactic category of the infinitive marker *å* in Norwegian is still debated in the literature, researchers are generally met with the same difficulties regarding the position of PRO. Some argue that PRO comes after the infinitive marker (Faarlund et al. 1997; Åfarli & Eide 2003; Faarlund 2007), while others claim that PRO is indeed positioned before the infinitive marker (cf. Faarlund, 2015).

Without any directly observable consequences, other than where the PRO is represented in surface structure, it is difficult to determine which of these structures is the more realistic depiction. However, in the case of Obligatory Control sentences, PRO is obligatorily coreferent with either the subject or the object of the preceding clause (Williams, 1980). There is psycholinguistic evidence (cf. Hestvik et al., 2010) from trace activation that priming effects can be reliably detected for a moved constituent at trace positions. This may also be true for PRO positions. Will this antecedent activation, if any, be at the position marked in (2) or (3)? That is, will a facilitation effect be detected before or after the infinitive marker? This is a clear, testable hypothesis.

Control sentences are very similar to Raising sentences at the surface level. Raising sentences involve syntactic movement of an NP trace, which predicts trace activation effects like those associated with typical NP traces. PRO is so similar to an NP trace that it has been suggested that Control sentences could be explained by syntactic movement (Boeckx et al., 2010; Hornstein, 1999; Manzini & Roussou, 2000). Hornstein (1999) presents a PRO-less theory of Control that is compatible with minimalist principles and the strict definition of Inclusiveness by Neeleman and van de Koot (2002). It states that the non-finite verb is the anaphoric item that is bound by the NP in the matrix clause, and consequently PRO does not exist (cf. Janke, 2003, 2008).

The experimental work in this chapter is motivated by a need for empirical evidence for or against the existence of PRO as a covert element. Previous research on NP traces (Nicol, 1988; Nicol & Osterhout, 1988; Nicol & Swinney, 1989; Hestvik et al., 2005, 2010; inter alia) has provided consistent results for trace reactivation. Nevertheless, results for PRO have been inconclusive (cf. Nicol, 1988; Nicol & Osterhout, 1988; Nicol & Swinney, 1989; Walenski, 2002). We seek to contribute to previous research on the processing of Control with data obtained on this phenomenon in Norwegian.

We predict that all instances of Obligatory Control (Subject and Object) will produce similar antecedent reactivation patterns if they are processed using the same mechanisms. If there is a difference between these two types of Control, there could be an additional mechanism or rule in place for coreference assignment in Control constructions. If only the correct antecedent shows evidence of reactivation then PRO will reflect NP trace characteristics, although the position of reactivation must also be taken into consideration. If both possible antecedents show evidence of reactivation then PRO will display pronoun characteristics (Nicol & Swinney, 1989).

Our experimental design investigates the position where reactivation can be found, comparing reactivation due to syntactic coreference with that of previous occurrence in the sentence (i.e. *non-syntactic reactivation*). In our main experiment, we establish a more precise baseline under laboratory conditions, which allows us to separate syntactic priming effects from non-syntactic effects of repetition.

1.1 Control sentences

Control clauses are present in every language, and, similarly, PRO is hypothesized to exist in all languages. Theoretically speaking, since PRO is thought to be part of Universal Grammar, its properties should be language independent, and therefore it is important to test this assumption across languages. Some evidence supporting the claim that the characteristics of PRO vary across languages has already been uncovered (Sigurðsson, 2008; Hornstein, 2014). Though our research is on the Norwegian language, we will provide examples for each sentence type in English to ease the burden on the reader.

1.1.1 *Subject Control*

The example in (4) is a case of Subject Control, as the subject of the matrix clause (*the elephant*) is the antecedent of PRO, the subject of the non-finite clause.

(4) [The elephant]$_i$ promised the alligator PRO$_i$ to (PRO$_i$) go to the river.

Example (4) illustrates the need for PRO and indicates the two positions in which it could theoretically be found. The sentence contains two verbs: *promised* within the matrix clause and *go* within the subordinate clause. The verb *promised* has two theta roles to distribute. These are given to the subject (*the elephant*) and the object (*the alligator*) of the matrix clause. The verb *go* also has a theta role to distribute, but the subordinate clause is missing its overt subject. *The elephant* is the actual subject of the subordinate clause, as it is the one that is going to go to the river, not *the alligator*. It is represented in the matrix clause, but, since it has already been assigned a theta role from the matrix clause verb, it cannot receive a theta role from the subordinate clause verb. PRO solves this distribution problem for the Theta Criterion, and Subject Control is established by a coreference relationship. This relationship is hypothesized to achieve reactivation of the antecedent at the PRO position.

1.1.2 *Object Control*

Example (5) below is an example of Object Control, as the object of the matrix clause (*the monkey*) is the antecedent of PRO.

(5) The hippo told [the monkey]$_i$ PRO$_i$ to (PRO$_i$) go to the river.

Here the subject of the non-finite clause is *the monkey*, who is the one that will go to the river. PRO takes the position of the subject of the non-finite clause and establishes a coreference relation with the object of the matrix clause, as a result of the Control verb in the matrix clause. One question is whether an observed effect for Object Control is indeed a *syntactic reactivation effect* or simply a *recency effect*. These can be teased apart using an appropriate baseline during data analysis.

1.2 Raising sentences

Some verbs have syntactic arguments that are not semantic arguments. One such verb is *seem*, whose subject does not receive a theta role from the verb *seem*. Raising sentences can be analyzed as the result of movement. An element, base-generated in the non-finite clause, moves into a syntactic position in the matrix clause, leaving a trace in its original position. Verbs like *assume* may similarly have a filler in their object role. Thus, in English, both subject and object positions may serve as a landing site for an argument from another predicate.

Raising sentences may look very similar to Control sentences, but they use matrix clause verbs that do not assign theta roles to all their arguments. This allows for NP-movement to positions in the matrix clause. Raising sentences have been argued to be a clear case of movement since Postal (1974).

Some verbs that have Exceptional Case Marking (ECM) make the subject of the non-finite subordinate clause appear with object Case in the matrix clause. These sentences are also called Subject-to-Object Raising. See example (6), where Adam might not believe *her* but rather the proposition that *she* is honest.

(6) Adam believes her to be honest.

Subject-to-Object Raising is virtually non-existent in Norwegian, which creates some challenges for matching Raising sentences with Object Control sentences. Norwegian also contains further syntactic constraints for Object Control, such as the Direct Object Constraint (DOC) (Lødrup, 2008) and specific animacy requirements. Additionally, it was not possible to use transitive verbs in the Subject-to-Subject Raising condition described in the next section. For these reasons, we will use Raising sentences mainly as filler sentences in the analysis presented in this article.

The intense difficulty that we encountered while creating the sentence stimuli for our second experiment supports a distinction between Control and Raising constructions. If these two sentence types have an identical syntactic structure, we should not have struggled to match minimal pairs across sentence conditions. NP traces and PRO are structural units that are theoretically assumed to have similar properties across all languages. Hornstein (1999), using data from English, suggests that Obligatory Control and Raising both contain NP traces and that PRO is not necessary in these Control constructions. However, our Norwegian data illustrates the importance of maintaining them as two separate syntactic constructions. In Norwegian in particular, Control and Raising seem to be subject to different constraints.

1.2.1 *Subject-to-Subject raising*

Example (7) shows a sentence where the subject of the non-finite clause has been raised to the subject position of the matrix clause.

(7) [The alligator]$_i$ seems [t_i] to catch a monkey every day.

The verb *catch* selects its subject (*the alligator*) and its object (*a monkey*). This structure can be combined with a predicate like *seem* that allows the base-generated subject (*the alligator*) to move into its subject position, leaving a non-finite clause as its complement. A trace resulting from syntactic movement is predicted at the position marked by t_i.

1.2.2 *Subject-to-Object raising*

Example (8) shows a sentence where the subject of the non-finite clause has been raised to the object position of the matrix clause.

(8) The monkey assumes [the banana]$_i$ [t_i] to be good for dinner.

The verb *be* selects its subject (*the banana*) and its complement. This structure can be combined with a predicate like *assume* that allows the subject (*the banana*) to move into its object position, leaving a non-finite clause as its complement. A trace is predicted at the position marked by t_i.

1.3 Our objective

We investigate Control in Norwegian, as most prior research on Control has been completed using English. We will present one pilot study and one main study. Our aim is to compare coreference processing and positioning of referent reactivation during the processing of the aforementioned sentence construction types. How do they compare to previous research on NP trace and pronoun processing?

2. The current study

Reaction time and priming experiments have been shown to be effective in the research of a variety of trace items. This can be seen, for example, in the studies discussed by Nicol and Swinney (1989) and Hestvik et al. (2010). The pilot study is motivated by the theoretical uncertainty surrounding the location of PRO. Are there processing effects present for PRO during online processing? If so, do they appear before or after the infinitive marker?

2.1 Objective and hypothesis: Pilot experiment

In the pilot experiment, we investigate the possibility of reactivation at PRO positions. It is assumed that PRO is reactivated either at position 1, before the infinitive marker, or position 2, after the infinitive marker. See examples (2) to (5) for an illustration. A priming effect related to PRO is thus noted if a significantly faster RT is recorded in one of these positions in comparison to the other, and specifically for the NP that is coreferent with PRO. The other matrix clause NP may show repetition priming.

The subject is activated from a longer distance than the object, the latter being very close to both test positions. Thus it can be expected that it will be harder to find a clear reactivation effect for the object, as this NP will be stored in more recent memory. The null hypothesis is that we will not see any significant differences for either position related to PRO.

2.1.1 *Rationale*

Considering the extensive research already completed on NP trace processing, it seems reasonable for the next step in research to include null items, such as PRO. The plan for our research was to contribute to foundational research on understanding Control processing.

2.1.2 *Participants*

For the first experiment, we recruited 22 native speakers of Norwegian at the University of Bergen. No participant information such as age or gender was recorded; however, all participants were around the average student age (between 18 and 30 years old). We are obliged to collect only minimal personal information about our participants according to the ethics guidelines at our lab. This applies in particular when participants are not compensated. Therefore, the guidelines for preventing harm and protecting the anonymity of participants have priority. Participation was anonymous, voluntary, and unpaid. We performed both an outlier procedure and a task compliance procedure to ensure that the task was understood and processed in a similar manner. The experiment took approximately 20 minutes per participant.

2.1.3 *Materials*

Data collection was conducted over two weeks in a soundproof room equipped with a Macintosh Mini computer with a UHD 4K high resolution screen and a Cedrus RB-540 response box. The experiment was programmed using Superlab 5. The stimuli sequence was randomized, with a new randomization for each subject.

The experiment consisted of 12 sentence templates. These included four Subject Control, four Object Control, and four filler templates, with four different verbs in each condition. The filler sentences were made up of unrelated structures that did not contain PRO.

Examples (9), (10), and (11) display one example of each test sentence type.

(9) Subject Control:
Alligatoren$_i$ *lovet sjiraffen PRO*$_i$ *å bade i sjøen snart*.
The alligator promised the giraffe to swim in the lake soon.

(10) Object Control:
Elefanten anbefalte sjiraffen$_i$ *PRO*$_i$ *å ta et bad før sengetid*.
The elephant recommended the giraffe to take a bath before bedtime.

(11) Filler sentence:
Flodhesten gikk med elefanten til konserten i går.
The hippo went with the elephant to the concert yesterday.

Each target sentence type is of the form NP1 VP NP2 å VP$_{inf}$. NP1 and NP2 were chosen from a set of four different common exotic animals: giraffe, hippo, elephant, or alligator. All 16 NP combinations were tested for each sentence. Each of the animal NPs had one image associated with it. We selected the image stimuli and animal NPs used in the experiment from the same category of common exotic animals, and they all had approximately the same word length in Norwegian. Figure 1 displays the image stimuli used for the experiment. We opted to have simple images in black, white, and gray, which eliminated possible effects of color. All of the images were chosen to be distinct and easily recognizable.

Figure 1. Pilot Experiment Image Stimuli

The experimental design also included four unrelated training sentences. Example (12) shows a training sentence used in the pilot experiment.

(12) Training sentence:
Flodhesten sprutet på alligatoren med vann fra innsjøen.
The hippo squirted the alligator with water from the lake.

We addressed the possibility of individual NP effects on variance in the statistical analysis by including the images as a random factor. In the analysis, we used a different intercept (starting point) for each image in order to control any remaining effects of the images as such. The images were not otherwise associated with any of the experimental conditions, but rather all images occurred equally in all experimental conditions. We also included a different starting point for each verb (sentence), which was used to control variance due to the verb.

2.1.4 *Procedure*

The pilot experiment used a picture priming paradigm, collecting reaction times for recognizing image stimuli. Our experimental design was based on previous research done on NP trace reactivation in gap positions of relative clauses by Hestvik et al. (2010). We let the subject read words, on a high resolution (4K) screen and then presented images at controlled sentence positions. The participants decided whether the image was previously mentioned in the sentence or not. We thought using images as stimuli would simplify the task and reduce the need of additional linguistic mechanisms when processing the stimuli, hence reducing the number of overall mechanisms necessary to complete the task.

After the instructions were presented, the experiment began with the presentation of training sentences. Each sentence appeared two times with two different test positions. The training sentences were repeated if answered incorrectly and the participant received written feedback. These sentences helped to get the participant familiar with the task and ensured good task compliance.

The sentence stimuli were presented word by word at a fixed rate of 500ms per word. An image was presented at the position before or after the infinitive marker, or at the end of the sentence. The image could be coreferent with either the subject or the object in the matrix clause, or it could be unrelated. The subjects were asked to answer whether the image had been mentioned previously in the sentence or not by pressing either a green or a red button. If the image had been named in the sentence, the expected answer was a *yes*. All other images were expected to have *no* answers.

The number of expected *yes* and *no* answers were balanced so that there would not be any obvious default answer. The number of correct answers were recorded and used to estimate task compliance as well as to investigate effects on accuracy related to antecedent reactivation. Participants were required to have more than 80 percent correct responses to be included in the analysis.

The experimental design contained three levels for test position: before or after the infinitive marker, and a position at the end of the sentence to account for possible delayed antecedent reactivation. There were two levels for sentence type: Object Control and Subject Control. There were two levels for priming: syntactic coreference with PRO or no coreference with PRO. The priming effect of PRO coreference is the difference between no coreference and coreference for each combination of Control type and position, calculated separately. An illustration explaining the levels of priming is included in Figure 2. The figure also shows two of the test positions used in the experiment, before and after the infinitive marker.

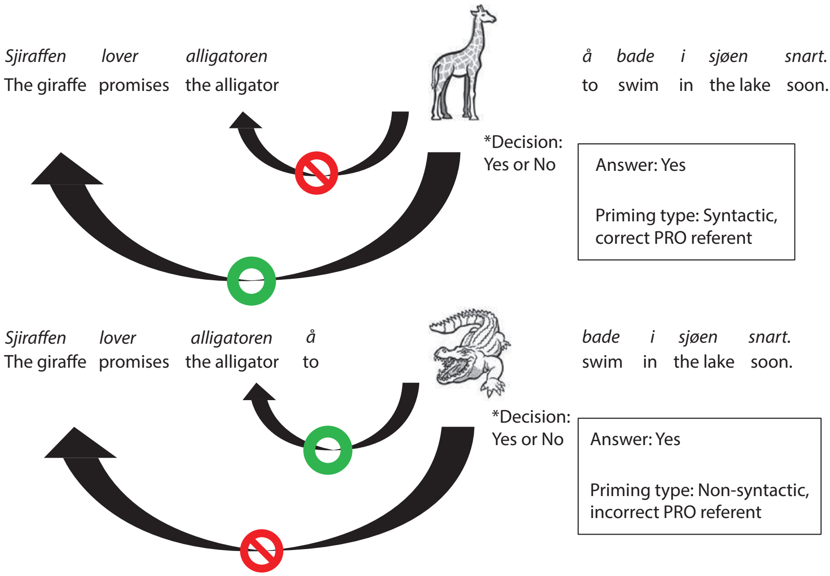

Figure 2. An illustration of the priming levels used in the experimental design

A circle is shown on the arrow when the image stimulus matches a previously mentioned NP. A no symbol (circle-backslash symbol) is shown on the arrow when the image stimulus does not match a previously mentioned NP. Both examples induce a *yes* answer from the participant but access different levels of priming. Both examples are of Subject Control. Depending on the grammatical role of the NP matching the image stimuli, the image represents reactivation due to syntactic coreference (*giraffe*) or repetition (*alligator*).

2.1.5 *Results*

Observations were analyzed with mixed effects models using the lmerTest package in R (Kuznetsova et al., 2016). These models allowed us to specify fixed and random effects. The fixed effects were Condition (i.e. type of Control (Subject Control, Object Control)), type of Priming (syntactic or repetition), and the test position (before/after the infinitive marker, or at the end of the sentence).

In the pilot study, we unintentionally introduced an unaccusative causative construction into the Subject Control condition. The formal subject is still valid, but it is a marked construction, as the formal subject is not the active agent. The sentence stimulus in question is shown in (13). To understand how we have categorized this sentence, compare it to the examples in (14) and (15). We can see from (14) and (15) that *learn* can appear as an intransitive verb, with or without an optional non-finite clause argument. In both of these examples, *learn* also appears as an unaccusative verb, as the grammatical subject is not the semantic agent.

(13) Sjiraffen lærte av elefanten å drikke fra elva.
The giraffe learned from the elephant to drink from the river.

(14) Sjiraffen lærte mye.
The giraffe learned a lot.

(15) Sjiraffen lærte å drikke fra elva.
The giraffe learned to drink from the river.

Now, looking back at (13), we see that it is the transitive counterpart of (15) using the same verb, *learn. Learn* can alternate between transitive and intransitive as an alternating ambitransitive verb (Dixon and Aikhenvald, 2000). In addition, it seems that *learn* maintains an unaccusative distribution of semantic roles whether it is transitive or intransitive. Note that transitive *learn* is a causative construction. As a consequence of the unusual distribution of semantic roles, the object is forced to be the causer and the subject to be the causee. Another characteristic of transitive *learn* is that it allows for underspecified object deletion (UOD) (cf. Alsina, 1992). Therefore, as is the case in (13), the semantic agent of *learn* is an oblique syntactic object that can be deleted and is not included in the argument structure of the verb.

We completed two statistical analyses, including and excluding all data from this sentence, to investigate whether this construction introduces additional variance into the dataset or drives any particular effects. We found no effect stemming from the inclusion of this sentence in our analysis. The calculated effects remained within milliseconds of each other; therefore, we will present the analysis that includes data from the unaccusative causative sentence stimulus.

Our analysis is based on all answers correctly identifying that an image was mentioned in word form in the presented sentence up to the point of image presentation. We have restricted the presentation to the two positions that are of most interest. More details are available in Larsen (2017). We estimate main and interaction effects. Significance of the factors are given by an Analysis of Variance (ANOVA) of the variance estimates delivered by the Mixed Effects Model (lmerTest), which also estimates the degrees of freedom for within-variance. The degrees of freedom can be thought of as the number of independent data points estimated by the model; thus, the model compensates for correlation between factors. The null hypothesis is that everything is equal, i.e. no effect of type of Control (Condition), Position, or Priming, nor their various interactions.

We used the formula in (16) to specify the model. Reaction times were analyzed as a function of Control Condition, Position, and type of Priming. We were interested in interaction effects between our fixed factors. Variance is controlled by the random factors. We assumed that each participant has their own baseline for reaction times and that they may react differently to each Control Condition. We also assumed that each verb may have different effects on Priming and that images may have differences in how easy they are to recognize.

(16) RT ~ Condition*Position*Priming + (Condition|Participant) + (Priming|Verb) + (1|Image)

There were a total of 2880 data points from 20 different participants. Two of the original 22 participants were excluded on the basis of poor task compliance. From the 2880 measurements, 169 were excluded due to wrong answers and three were excluded due to no response within a predefined 3000ms interval.

The type III Analysis of Variance (ANOVA) with Satterthwaite approximations for the model showed significant effects for the interaction between Control type and Priming (F (1, 540.40) = 13.25, $p < 0.0003$ ***) and Control type, Position, and Priming (F (1, 541.71) = 6.13, $p < 0.014$ *). Thus we detected effects related to antecedent reactivation. These effects depend on the type of Control as well as priming and sentential position; therefore, we can reject the null hypothesis and take a closer look at the significant interaction effects.

When we look at the difference of effect between position 1 and 2, we see that for Object Control sentences there is a difference of 127.67 – 29.13 = 98.54 (i.e. a slowdown, or interference). For Subject Control sentences, there is a difference of – 73.87 – (– 9.31) = – 64.56 (i.e., a speed-up or facilitation). Subject Control paired with the correct syntactic antecedent has faster recognition in position 2 than in position 1, and the opposite was observed for Object Control. This suggests antecedent reactivation indeed takes place in Control constructions but that reactivation functions differently based on Control type and position.

For Subject Control, position 2 is the position that is furthest away from the antecedent word. However, as we observed, syntactic referent reactivation for Subject Control was fastest in the second position, right after the infinitive marker. This is the position in which PRO is thought to be generated before any movement.

From the significance testing, we noted differences for syntactic priming of a PRO antecedent word. The results for Object Control may be explained by recency priming. The Subject Control sentences have more of a reactivation effect at position 2, which cannot be explained by recency effects. These results encourage an analysis that supports antecedent reactivation after the infinitive marker in a Control sentence.

2.1.6 *Summary pilot study*

We found significant effects related to a possible activation of PRO. In Subject Control, the correct referent of PRO had faster responses in position 2 (after the infinitive marker). For Object Control there was a somewhat unexpected slowdown at position 2, which may indicate competition at that position.

2.2 Objective and hypotheses: Main experiment

From the pilot experiment, we learned that position 2 is a strong candidate for PRO activation. In the main experiment, we want to more closely control the effect of repetition priming. We compare words that have an equal distance and the same functional role by comparing NPs syntactically coreferent with PRO with the same word mentioned in the same role but in the other Control condition. Furthermore, we want to test if PRO is activated for coreference in both Subject and Object Control. Due to recency effects, we are aware that it may be harder to detect activation for Object Control. The null hypothesis is that we will not find significant differences.

2.2.1 *Rationale*

From the results provided by the pilot experiment, it was evident that the experimental methods used were capable of detecting PRO processing effects. We considered the idea that the size of the experiment (number of participants) and the amount of data collected from the pilot experiment were not enough for a thorough statistical analysis.

We decided to continue with an almost identical experimental design for the main experiment. A few small changes were made, including an extra test sentence and image stimuli. An important lesson learned from the pilot experiment was where to focus our test positions. The three positions utilized in the pilot experiment created too much variance in our statistical calculations.

2.2.2 *Participants*

For the second, larger experiment, we recruited 64 native speakers of Norwegian, 32 male and 32 female, from the same general area as those recruited for the pilot experiment. Again, no participant information such as age or gender was recorded in compliance with our ethical guidelines. The majority of participants were around the average student age (between 18 and 30 years old), with a few being between 50 and 60 years old. Age was not considered an important factor in our experimental design. The participants were required to have Norwegian as a native language, and an outlier procedure was completed during the data analysis. As before, all participation was anonymous, voluntary, and unpaid. The length of the experiment was approximately 15 minutes per participant.

2.2.3 *Materials*

Data collection took place over a period of two months, and we used the same equipment that was used for the pilot experiment.

The experiment utilized 20 sentence templates, containing four sentences of each of the five sentence construction types. Two sentence types were relevant for the study we present here, namely Subject Control and Object Control. The other sentence types, Raising and Intransitive Subject Control, are considered filler sentences. This made the variation of sentence constructions in the experiment more realistic. It was more difficult for our subjects to form conscious problem-solving strategies. It also minimized the risk that versions of a sentence we are interested in were presented close to each other. Each sentence was constructed with a different verb across conditions. We give one example of each sentence type in (17) to (21).

(17) Transitive Subject Control:
Alligatoren$_i$ *lover sjiraffen PRO*$_i$ *å bade i sjøen snart.*
The alligator promises the giraffe to swim in the lake soon.

(18) Intransitive Subject Control:
Apekatten$_i$ *lærer PRO*$_i$ *å sprute vann ut av nesen.*
The monkey learns to squirt water out of the nose.

(19) Object Control:
Elefanten tillater sjiraffen$_i$ *PRO*$_i$ *å plukke blomster i hagen.*
The elephant allows the giraffe to pick flowers in the garden.

(20) Subject-to-Subject Raising:
Alligatoren$_i$ *synes t*$_i$ *å fange fisk til frokost hver dag.*
The alligator seems to catch fish for breakfast every day.

(21) Subject-to-Object Raising:
Bananen$_i$ antar apekatten t_i å være god til middag.
The banana$_i$ assumes the monkey, to be good for dinner.

The sentence types were constructed similarly to those in the pilot experiment with the form NP1 VP (NP2) å VP$_{inf}$. Four NPs were selected from the set of common exotic animals: giraffe, elephant, alligator, and monkey. We exchanged the hippo for the monkey because of its likeness to the elephant. Each word was associated with one image, and their frequency of appearance was balanced throughout the experiment.

Four other NPs were included in the experiment in order to satisfy the animacy requirements for Subject-to-Object Raising constructions in Norwegian. These inanimate NPs were only used in this Raising condition, and the frequency of each NP appearance was balanced throughout this condition.

We selected the images using the same constraints and qualifications that were used in the pilot experiment. Similarly, we continued to use a different intercept for each image to control effects stemming from image choice. The image stimuli utilized for the main experiment are shown in Figure 3. The inclusion of more images also made the task more realistic for targeting reactivation, rather than just familiarity. The possibility of correctly guessing whether an image was mentioned goes from two out of four to two out of eight.

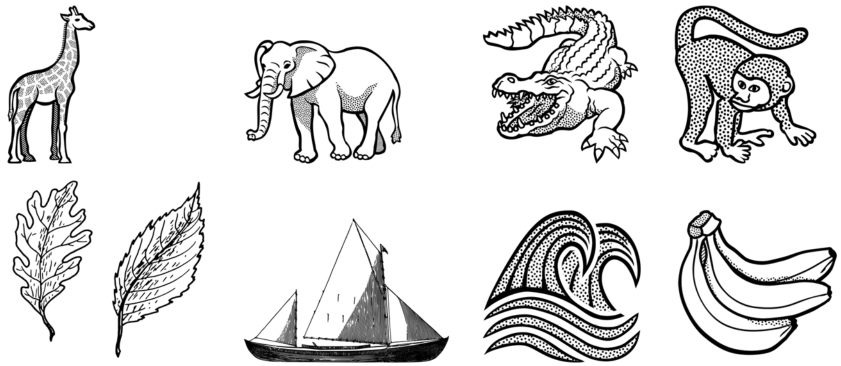

Figure 3. Main Experiment Image Stimuli

The experimental design of the main experiment also used training sentences with unrelated structures as in the pilot experiment. These training sentences functioned in the same way and were included for the same reasons; however, twelve sentence presentations were present in the main experiment in comparison to the eight used in the original pilot experiment. We included a separate intercept for each verb, as in the pilot experiment data analysis.

2.2.4 *Procedure*

We created the main experiment with a similar design to that of the pilot experiment. The main differences included a differing number of test positions and sentence types. There were also two comprehension questions asked to each participant after the conclusion of the experiment, which were not included in the pilot experiment. The answers did not affect the results, but the participants were warned ahead of time about the questions in order to boost task compliance and attention.

This experiment was created using Superlab 5 and contained two separate experimental blocks, with only one block or version shown per participant in order to counterbalance the presentation order for position one and two. The blocked design was also used to limit the length of the experiment to avoid fatigue effects.

The experiment followed the same procedure as the pilot and the stimuli sequence was randomized in the same manner. After the participant read and understood the instructions, the training sentences were presented, along with written feedback to ensure task comprehension. The experiment then transitioned to test sentence presentations without feedback.

The experimental design was comprised of two levels for test position: before or after the infinitive marker. There were five levels for sentence type, but only two are included in our data analysis: transitive Subject Control and Object Control. Finally, there were two levels of priming: syntactic coreference with PRO and no coreference with PRO (repetition). For an illustration and short explanation of the priming levels, see Figure 2.

2.2.5 *Results*

We completed the statistical analysis for the main experiment using mixed effects models with specified fixed and random effects, similarly to the pilot. The fixed effects were: type of Priming (syntactic or repetition) and the test Position (before/after the infinitive marker).

We improved the baseline such that for Subject Control the baseline is the subject of Object Control, and vice versa the baseline for Object Control is the object of Subject Control. The baseline is for words at the same distance, and with the same functional role. It is only the verb of the sentences that are different at the point of testing, which is before or after the infinitive marker. The subject is further away from the point of testing, and the object is very close to the point of testing, but we are comparing effects with only minimal differences, including reactivation of PRO (and the verb controlling PRO, which will not be possible to separate as the verbs are unique to the condition of Subject or Object Control).

We maintain the null hypothesis that all effects are equal: there is no effect of any fixed factor or interaction of fixed factors. Since we pair one sentence type condition with the other to create the baseline, we have to separate the analysis into two tests: one for Subject Control and one for Object Control, in order to ascertain that data points are independent.

One sentence item was excluded from the analysis. We unintentionally introduced a passive construction into the Subject Control condition. The sentence stimulus in question is shown in (22). The object (and patient) of the verb *ask* is the alligator, and it has been promoted to the subject position of the sentence. Meanwhile, the original subject of the verb *ask*, the one doing the *asking*, has be demoted and reintroduced as an oblique object.

(22) Alligatoren blir bedt av apekatten å være en god venn.
The alligator is asked by the monkey to be a good friend.

We completed two statistical analyses, including and excluding all data from this sentence. After witnessing that this passive stimulus affected the significance of priming in one condition, we chose to present the analysis that excludes data from the passive stimulus, as to prevent any effects being driven by this alternative construction. Only data with the correct affirmative answer were included in the analysis.

One possible problem for our analysis is that the object position is always closer to the target test position. Thus we expect that it will be more difficult to ascertain effects for Object Control in the current design. In example (23), *elefanten* involves only repetition priming if an image of it is shown at position P_1 or P_2. On the other hand, *apekatten* is coreferent with PRO and therefore involves both syntactic as well as repetition priming. A similar occurrence takes place in example (24), but the situation is reversed for the two words. Consequently, there is a chance that the recency of the object will affect reaction times, not only the reactivation of a PRO referent. This is another motivation for analyzing Subject Control and Object Control separately.

(23) Subject Control:
Apekatten lover elefanten [P_1] å [P_2] bade i sjøen.
The monkey promises the elephant to swim in the lake.

(24) Object Control:
Apekatten nekter elefanten [P_1] å [P_2] hente mat.
The monkey forbids the elephant to fetch food.

If it is assumed that PRO occupies only one syntactic position then position also defines a possible priming effect. However, we do not exclude the possibility that effects of PRO can be detected at either of the two positions, due to the possible movement of PRO. We are interested in the differences in priming effects between the two positions.

We recorded a total of 5952 data points from 64 different participants. Two participants were excluded due to low task compliance, and one more fell out after excluding passives. In total, 218 data points were excluded due to wrong answers, and six were excluded due to no response within a pre-defined 3000ms interval. The entire experiment included Subject Control with transitive verbs, Subject Control with intransitive verbs, Object Control, Subject-to-Subject Raising, and Subject-to-Object Raising. After removing outliers (RTs restricted to between 406ms and 1275ms), we have 418 observations of subject NP presentations and 461 observations of object NP presentations, distributed over 8 verbs. Then excluding the possibly problematic passive gives a data set with 7 verbs, and 361 observations for subject and 401 for object presentations.

A dataset can be analyzed in many ways. We used a novel baseline to compare the instances where the same NPs are in the same sentence positions across Control conditions. We matched repetition-only with instances of syntactic reactivation, as syntactic priming also included the effect of simply having seen the word before (repetition). Examples (25) and (26) show one pairing. The NPs maintain the same distance from PRO and the only difference before the PRO position is the Control verb. This is controlled in the analysis by having a separate intercept for each verb.

(25) Subject Control:
Sjiraffen tilbyr alligatoren å ta ansvar for sine feil.
The giraffe offered the alligator to take responsibility for his mistakes.

(26) Object Control:
Sjiraffen fraråder alligatoren å spise mat for sent.
The giraffe warns the alligator (not) to eat food too late.

This baseline *directly addressed the uncertainty factor* related to memory activation effects at various linear distances. We have included interaction plots to illustrate the analysis results using this baseline. Plots for both Control conditions are shown in Figure 4. As we can see in the plot on the left, Subject Control shows a significant effect for Position. Syntactic Priming for Subject Control in Position 2 is 55.16ms faster than the intercept (Position 1 and repetition Priming). These results again support antecedent reactivation after the infinitive marker in Subject Control sentences.

We see a small, but not significant, effect of Position in the opposite direction for Object Control and a syntactic Priming effect approaching significance. Syntactic Priming for Object Control in Position 2 is 25.24 ms *slower* than the intercept (Position 1 and repetition Priming). Though the effect of Position is not shown to be significant, it is interesting to notice that the direction of the effect for Object Control, which tends towards interference rather than facilitation, is

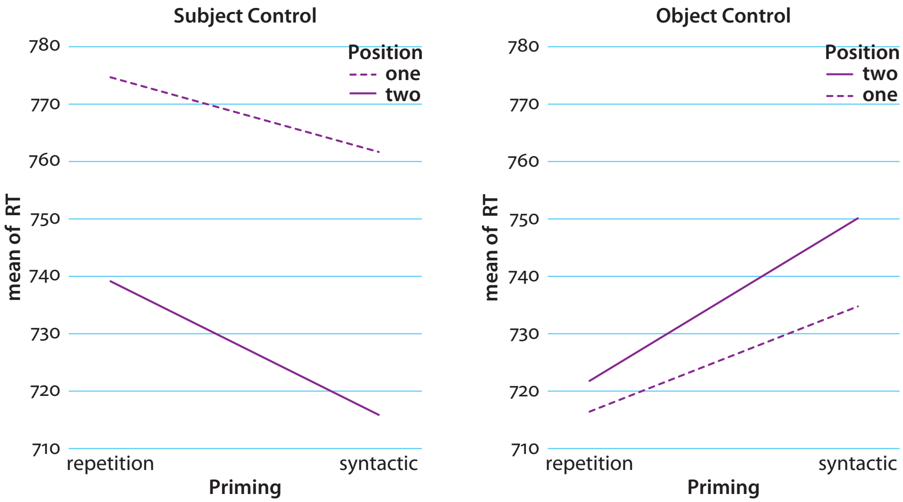

Figure 4. Priming Effects for Subject Control and Object Control Using a Matched Baseline

similar to the pilot study. One possibility would be that reactivation of the object as the PRO referent is in competition with the reactivation of the subject. This is congruent with always activating the subject at position two, where activation of the object is slowest. More data is needed to confirm if there is indeed competition between subject and object. These results, in addition to those recorded for the pilot study, indicate that Subject and Object Control function differently regarding PRO antecedent reactivation. PRO referent reactivation at position 2 is consistently faster for Subject Control but not for Object Control.

The interaction plots and data both indicate that Object Control experiences a *slowdown effect* due to syntactic priming. The *opposite* occurs for Subject Control. It is possible that syntactic priming creates an interference effect for Object Control constructions if the subject is always (re)activated.

Table 1. Regression equations

Condition	Intercept	Pos2	Synt	Pos2&Synt
SubjC	777.19	−28.88	−8.40	−17.88
ObjC	722.14	+17.60	+13.73	−6.09

The type III ANOVA of the mixed effects models with Satterthwaite approximations show significant effects ($p < 0.05$) of Position for Subject Control: $F(1, 100.4) = 5.48$ ($p = 0.021$), with a large effect size (partial-$\eta = 82.5\%$, 84.6%), but no significant effect for Object Control. The regression equations used are included in Table 1.

If we focus on the effect of Position, the data suggests that position works in opposite ways for Subject Control and Object Control. Subject Control displays noticeable differences between position 1 and 2, with a significantly faster mean RT in position 2. Object Control shows no significant differences.

3. Discussion

3.1 Pilot experiment

The pilot experiment discovered significant effects related to Priming, Position, and type of Control. This allowed us to reject the null hypothesis which stated that no effects would be found for any single fixed effect or combination of fixed effects. For Subject Control, there was a significant effect specifically related to Position: Subject Control was associated with faster reaction times in position 2, after the infinitive marker, compared to position 1, before the infinitive marker. This was also replicated in the main experiment. Results for Object Control were consistent with the fact that objects in the matrix clause were closer to the test position where the probe picture was shown.

From the pilot experiment, it was clear that antecedent reactivation effects were different between Subject Control and Object Control. The results from the pilot experiment were promising and motivated the main experiment.

3.2 Main Experiment

The results of the main experiment reinforced our previous findings that we can detect effects related to antecedent reactivation during the on-line processing of Control sentences. We will discuss the results of each condition (sentence type) in separate subsections.

3.2.1 *Subject Control*

In Figure 4, we saw that reaction times at position 2 are faster than those at position 1 for both repetition priming and syntactic priming. This effect is significant at $p < 0.05$. It is also a fairly large effect, with an effect size (partial-eta) above 80%. See the appendix for the full ANOVA table. From the figure, we can also see that syntactic priming at both positions is faster than repetition priming alone, but this effect is not significant. This shows how hard it is to separate repetition and syntactic priming. This is consistent with the findings of previous research that found that both PRO and pronouns activated all possible NP candidates when assigning coreference (Nicol and Swinney, 1989).

3.2.2 *Object Control*

In Figure 4, we also saw that reaction times for Object Control and syntactic priming are *slower* in both positions. This effect did not reach significance; however, we should be careful about asserting that there is no effect of priming for Object Control, since it could be an issue of statistical power. If there is interference, what is the object competing against? One possible answer is that the subject gets reactivated and competes against the object during coreference assignment. Support for this suggestion is that the slowdown is larger at position 2, where the subject has been shown to be significantly reactivated in both the pilot and the main study.

Nonetheless, the effects for Object Control were not statistically significant and require further testing. This requires us to revise our original research question with respect to reactivation effect detection: Can antecedent reactivation effects be detected during on-line processing for both Subject *and* Object Control?

3.2.3 *Interpretation Subject Control and Object Control*

There are many reasons why priming effects and position effects were extraordinarily difficult to detect for Object Control in our experiments. First of all, the test positions are linearly close to the antecedent. In position 1, the antecedent is the word immediately before the picture presentation. In position 2, only an infinitive marker comes between the antecedent and the image stimuli presentation. The close proximity of the antecedent implies that there might be no need for reactivation, as it is already active in memory. It is possible that position 2 is where the object of the matrix clause meets the greatest competition during antecedent assignment. It is also possible that the effect is asymmetric: the furthest antecedent is also more strongly activated, whereas the closest antecedent will have to compete.

Our research question concerning the reactivation of the syntactic antecedent of PRO is at this point inconclusive. However, the evidence points to activation of all possible NP antecedents in the matrix clause directly following the infinitive marker. This is supported by the significant effect of position for both syntactic and repetition priming for Subject Control. It may be *conjectured* from the close to significant slowdown for syntactic priming at both positions for Object Control as well.

If the slowdown can be replicated, it would indicate that the activation of the object from the matrix clause is in competition with the reactivating matrix clause subject. Thus, we may speculate that assigning the object of the matrix clause as the antecedent of PRO requires some additional computation that involves accessing the Control information of the verb in the matrix clause. This would be in line with a popular pragmatic argument for (null) pronoun resolution in which the most prominent discourse entity is selected as the pronoun referent (cf. Schumacher, 2016). The parser initially selects for the matrix clause subject due to its prominent status and then must reanalyze and reassign the antecedent after receiving additional information from the argument structure of the verb. For repetition priming, there would only be activation from having seen the name of the picture before.

4. Conclusions

Our results point to reactivation of the matrix clause subject after the infinitive marker of a non-finite clause. Insignificant effects before the infinitive marker could be expected by anticipation of the infinitive marker. We argue that the subject may be reactivated in competition with the object, even in cases of Object Control, although this claim requires further support from experimental evidence. The availability of both referents points towards a similarity with ordinary pronoun resolution.

In order to better control the effect of linear distance, future research may control the distance between PRO and the object by, for example, adding modifying prepositional phrases in order to further remove the object from the point of decision. Similarly, the subject can be modified by an adjective to ensure both the matrix clause subject and object NPs maintain a comparable level of complexity. Even though more research is necessary to confirm the patterns of PRO reactivation, we believe that we have taken important steps towards understanding the nature of PRO in natural language processing.

Acknowledgement

We are grateful to our anonymous reviewers, who gave valuable critique that significantly increased the readability and clarity of argumentation in this article. We would also like to thank professor emeritus Robert James Jarvella for his comments and positive encouragement to publish after reading the penultimate manuscript, and professor Marit Julien, at Lund University, for her kind interest and suggestions for further reading.

References

Åfarli, T. A., & Eide, K. M. (2003). *Norsk generativ syntaks*. Oslo: Novus Forlag.

Alsina, A. (1992). On the argument structure of causatives. *Linguistic Inquiry*, 23(4), 517–555.

Baltin, M. (1995). Floating quantifiers, PRO and predication. *Linguistic Inquiry*, 26(2), 199–248.

Boeckx, C., Hornstein, N., & Nunes, J. (2010). *Control as movement*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511761997

Carnie, A. (2013). *Syntax: A generative introduction* (3rd ed.). Chichester: John Wiley & Sons.

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: The MIT Press.

Chomsky, N. (1993). *Lectures on government and binding: The Pisa lectures*. Berlin: Mouton de Gruyter. (Original work published 1981). https://doi.org/10.1515/9783110884166

Chomsky, N., & Lasnik, H. (1993). The theory of principles and parameters. In J. Jacobs, A. von Stechow, W. Sternefeld & T. Vennemann (Eds.), *Syntax: An international handbook of contemporary research* (Vol. 1, pp. 506–569). Berlin: Walter de Gruyter.

Dixon, R. M. W., & Aikhenvald, A. Y. (Eds.). (2000). *Changing valency: Case studies in transitivity*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627750

Faarlund, J. T. (2007). Parameterization and change in non-finite complementation. *Diachronica*, 24(1), 57–80. https://doi.org/10.1075/dia.24.1.04faa

Faarlund, J. T. (2015). The Norwegian infinitive marker. *Working Papers in Scandinavian Syntax*, 95, 1–10.

Faarlund, J. T., Lie, S., & Vannebo, K. I. (1997). *Norsk referansegrammatikk* (Norwegian reference grammar). Oslo: Universitetsforlaget.

Hestvik, A., Nordby, H., & Karlsen, G. (2005). Antecedent reactivation by surface and deep anaphora in Norwegian. *Scandinavian Journal of Psychology*, 46(3), 229–38. https://doi.org/10.1111/j.1467-9450.2005.00452.x

Hestvik, A., Schwartz, R. G., & Tornyova, L. (2010). Relative clause gap-filling in children with Specific Language Impairment. *Journal of Psycholinguistic Research*, 39(5), 443–456. https://doi.org/10.1007/s10936-010-9151-1

Hornstein, N. (1999). Movement and Control. *Linguistic Inquiry*, 30(1), 69– 96. https://doi.org/10.1162/002438999553968

Hornstein, N., & Nunes, J. (2014). Minimalism and Control. In A. Carnie, D. Siddiqi, & Y. Sato (Eds.), *Routledge handbook of syntax* (pp. 239–263). Abingdon: Routledge. https://doi.org/10.4324/9781315796604

Janke, V. (2003). A PRO-less theory of Control. In A. Neeleman & R. Vermeulen (Eds.), *University College London Working Papers in Linguistics (UCLWPL)* (Vol. 15, pp. 213–242). London: UCL.

Janke, V. (2008). Control without a subject. *Lingua*, 118(1), 82–118. https://doi.org/10.1016/j.lingua.2007.05.002

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). *lmerTest: Tests in linear mixed effects models*. R package version 2.0-32.

Larsen, T. N. (2017). *Control and raising: Gone with(out) a trace?* (Unpublished MA thesis). The University of Bergen.

Lødrup, H. (2008). Raising to object in Norwegian and the derived object constraint. *Studia Linguistica*, 62(2), 155–181. https://doi.org/10.1111/j.1467-9582.2008.00147.x

Manzini, M. R., & Roussou, A. (2000). A minimalist theory of A-movement and control. *Lingua*, 110(6), 409–447. https://doi.org/10.1016/S0024-3841(00)00006-1

Martin, R. (2001). Null case and the distribution of PRO. *Linguistic Inquiry*, 32(1), 141–166. https://doi.org/10.1162/002438901554612

Neeleman, A., & van de Koot, H. (2002). The configurational matrix. *Linguistic Inquiry*, 33(4), 529–574. https://doi.org/10.1007/978-1-4613-2727-1_15

Nicol, J. (1988). *Coreference processing during sentence comprehension* (Unpublished doctoral dissertation). MIT.

Nicol, J., & Osterhout, L. (1988). Reactivating antecedents of empty categories during parsing (Unpublished manuscript). University of Arizona, Tucson.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18(1), 5–19. https://doi.org/10.1007/BF01069043

Postal, P. (1974). *On raising: One rule of English grammar and its theoretical implications*. Cambridge, MA: The MIT Press.

Radford, A. (2004). *Minimalist syntax: Exploring the structure of English*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511811319

Reed, L. (2014). *Strengthening the PRO hypothesis*. Berlin: Mouton de Gruyter.

Rosenbaum, P. (1967). *The grammar of English predicate complement constructions*. Cambridge, MA: The MIT Press.

Schumacher, P. B., Dangl, M., & Uzun, E. (2016). Thematic role as prominence cue during pronoun resolution in German. In A. Holler & K. Suckow (Eds.), *Empirical perspectives on anaphora resolution* (pp. 121–147). Berlin: Walter de Gruyter.

Sigurðsson, H. (2008). The case of PRO. *Natural Language & Linguistic Theory*, 26(2), 403–450. https://doi.org/10.1007/s11049-008-9040-6

Walenski, M. (2002). *Relating parsers and grammars: On the structure and real-time comprehension of English infinitival complements* (Unpublished doctoral dissertation). University of California, San Diego.

Williams, E. (1980). Predication. *Linguistic Inquiry*, 11(1), 203–238.

Appendix

The appendix includes the ANOVA table for the main experiment analysis. The additional tables show the test sentence stimuli for the pilot experiment and the main experiment.

	SS	F(between, within)=F	Pr(>F)
Subject Control			
Priming	19662	F(1, 60.2) = 0.99	0.32
Position (η = 84.6)	108852	F(1, 100.4) = 5.48	0.021
Priming:Position	6868	F(1, 230.5) = 0.35	0.56
	SS	**F(between, within)=F**	**Pr(>F)**
Object Control			
Priming	17976	F(1, 153.2) = 0.86	0.35
Position	8570	F(1, 65.8) = 0.41	0.52
Priming:Position	834	F(1, 141.1) = 0.04	0.84

Pilot Study

Subject Control	Object Control
Sjiraffen tilbød alligatoren å finne mat.	Elefanten anbefalte flodhesten å ta et bad før sen-getid.
'The giraffe offered the alligator to find food.'	*'The elephant recommended the hippo to take a bath before bedtime.'*
(x) Sjiraffen lærte av elefanten å drikke fra elva.	Elefanten tillot sjiraffen å danse for barna.
'The giraffe learned from the elephant to drink from the river.'	*'The elephant allowed the giraffe to dance for the children.'*
Alligatoren lovet sjiraffen å bade i sjøen snart.	Flodhesten nektet elefanten å spise bananer til middag.
'The alligator promised the giraffe to swim in the lake soon.'	*'The hippo denied the elephant to eat bananas for dinner.'*
Alligatoren skyldte flodhesten å lage lunsj.	Flodhesten frarådet alligatoren å ga til festen.
'The alligator owed the hippo to make lunch.'	*'The hippo advised the alligator (not) to go to the party.'*
	**Note: This sentence does include negation but translates to "advise against". The negation is there to maintain the infinitive clause in the English translation.*
Filler sentences	
Flodhesten gikk med elefanten til konserten i går.	Elefantens venn, alligatoren, fikk en stor gave til jul.
'The hippo went with the elephant to the concert yesterday.'	*'The elephant's friend, the alligator, got a big present for Christmas.'*
Sjiraffen og flodhesten sprang et løp i kveld.	Alligatoren bet sjiraffen på nesen på mandag.
'The giraffe and the hippo ran a race tonight.'	*'The alligator bit the giraffe on the nose on Monday.'*

Main Study

Subject Control	Object Control
Apekatten lover elefanten å bade i sjøen snart.	Elefanten tillater sjiraffen å plukke blomster i hagen.
'The monkey promised the elephant to swim in the lake soon.'	*'The elephant allows the giraffe to pick flowers in the garden.'*
Sjiraffen tilbyr alligatoren å ta ansvar for sine feil.	Apekatten nekter elefanten å hente noe mat til middagen.

'The giraffe offers the alligator to take responsibility for his errors.'	*'The monkey denies the elephant to retrieve food for dinner.'*
Elefanten skylder sjiraffen å lage lunsj til dem begge.	Sjiraffen anbefaler alligatoren å ta et bad før sengetid.
'The elephant owes the giraffe to make lunch for both of them.'	*'The giraffe recommends the alligator to take a bath before bedtime.'*
(x) Alligatoren blir bedt av apekatten å være en god venn.	Alligatoren fraråder apekatten å spise mat for sent.
'The alligator is asked by the monkey to be a good friend.'	*'The alligator advises the monkey (not) to eat food too late.'*
	**Note: This sentence does include negation but translates to "advise against". The negation is there to maintain the infinitive clause in the English translation.*
Intransitive Subject Control	
Apekatten lærer å sprute vann ut av nesen.	Sjiraffen ønsker å gå hjem så snart som mulig.
'The monkey learns to squirt water out of the nose.'	*'The giraffe wishes to go home as soon as possible.'*
Elefanten klarer å finne mat til frokost.	Alligatoren forsøker å bli full på råtne bananer.
'The elephant manages to find food for breakfast.'	*'The alligator tries to become drunk on rotten bananas.'*
Subject-to-Subject Raising	**Subject-to-Object Raising**
Alligatoren synes å fange fisk til frokost hver dag.	Sjøen anser sjiraffen å ha for mange bølger.
'The alligator seems to catch fish for breakfast every day.'	*'The lake, thinks the giraffe, to have too many waves.'*
Apekatten tenderer å klatre i tre for moro skyld.	Bananen antar apekatten å være god til middag.
'The monkey tends to climb trees for fun.'	*'The banana, assumes the monkey, to be good for dinner.'*
Elefanten begynner å spise bananer til jul.	Blader erklærer elefanten å være veldig nyttig til tørking.
'The elephant begins to eat bananas for Christmas.'	*'Leaves, declares the elephant,to be useful for drying.'*
Sjiraffen slutter å invitere elefanten på middag.	Båten forstår alligatoren å være full av mat.
'The giraffe stops to invite (inviting) the elephant for dinner.'	*'The boat, understands the alligator, to be full of food.'*

Priming paradigmatic gaps

Joakim Vea and Christer Johansson
University of Bergen

We present an experimental study on paradigmatic gaps in the neuter gender of Norwegian adjectives. The formation of neuter gender forms is typically not complicated, and yet people hesitate to use a predictable set of adjectives. We investigate the effect of priming in order to determine if the missing neuter forms will facilitate lexical decisions for their common gender form, possibly through lemma activation. A clear difference was found between regular and problematic adjectives. This shows that the intuition of speakers is reflected in detectable differences in processing.

Keywords: masked priming, paradigmatic gaps, pseudowords

1. Introduction

Imagine that you could not say a simple thing like "There is one proud child" in English, just because the grammar demands that 'proud' should be marked for the same morphological class as 'child', but no such form of 'proud' exists. So you would have to rephrase yourself ('The child, she is proud.'). This is the situation in Norwegian and Swedish for many adjectives. Gender marking and gender congruency are parts of linguistic classification systems. Gender classes are often assigned seemingly arbitrarily, but often influenced by phonological and semantic patterns, even though the logic is not obvious. These patterns may at times help to predict words, but in other cases it may block options that we think ought to be available. Grammaticality makes sentences easier to process. This is a grand generalization that implies that when rules are broken there will be some price to pay in the flow of processing. This article investigates the phenomenon of paradigmatic gaps. These are typically forms of words that, for no obvious reason, do not normally occur in grammatical sentences.

Crucially, it is not because the word form is already taken, or in competition with other words, and it is not because there is an irregular form that blocks the

https://doi.org/10.1075/lald.64.06vea

expression of a regular form (as in *went/*goed*). The form is completely missing at the surface level. Is it possible to observe such conflicts in word formation rules using an experimental paradigm?

The specific paradigmatic gap under investigation is a quirk of adjective inflection that occurs in both Norwegian and Swedish, where the neuter singular indefinite forms of some adjectives are missing from the inflectional paradigm. Language users demonstrate the phenomenon by their avoidance of using the missing neuter gender forms and by their hesitation in recognizing these forms as correct in general.

Paradigms that lack some expected forms, or have unexpected and systematic restrictions on their usage, are known as defective paradigms. We refer to the forms missing from the paradigm as a paradigmatic gap. The phenomenon of paradigmatic gaps is found in other languages as well (cf. Baerman, Corbett & Brown, 2010). Examples from English include a small number of verbs, such as 'beware' or 'quoth', that can no longer inflect regularly and some adjectives that only work in the predicative ('the cat is alive' vs ? 'the alive cat'). Speakers of English have acquired information that also prevents them from using forms such as 'adead'. Yang (2015) suggests a distributional solution to how learners may acquire knowledge of paradigmatic gaps for such *a*-adjectives (e.g., *alive/adead*) by observing the lack of (statistically) expected observations. Distributional approaches for similar problems (for example, the distribution of the dative alternation) have been presented previously (Jenset & Johansson, 2013; Johansson, 2017); Johansson (2017) gives an elaborate procedure for evaluating whether a pattern is over-, or under-, represented in contrast to a baseline.

The defective Norwegian paradigm discussed in the present article is more difficult to model than English *a*-adjectives, as *a*-adjectives, and their distributions, are easily recognized by their form. Our missing neuters appear with adjectives that are difficult to identify as problematic unless both form and semantics are considered. This is discussed below in the section on the morphology of problematic adjectives.

Paradigmatic gaps are of general interest from a linguistic perspective. They are relevant to studies of how we learn lexemes. For many low-frequency words, it is unlikely that we just store all forms as we encounter them. Rather, we use a word-formation regularity when we need it. This is at the center of Baker's (1979) Paradox and the problem of Learnability (Pinker, 1989). The problem of learnability is a lack of negative evidence that would make it impossible to correct an *over-generalization*. This triggered a long discussion, where connectionist and cue-based theories provided alternatives. We will not go into this discussion, but note that the paradigmatic gaps show a different, but related, problem of *under-generalization*: what stops the generalization of a simple rule when there are no alternatives? One possibility is that there is a productive process at work, blocking a regular alternative from activating a lemma. Our aim is to see if we can observe such a process.

Some previous research has involved the Norwegian, or more frequently the Swedish, defective adjective paradigms (cf. Johansson, 1999, 2003; Johansson & Torkildsen, 2005; Pettersson, 1990; Raffelsiefen, 2004; Rice, 2007; Fanselow & Féry, 2002; Lowenadler, 2010). While our experiment specifically uses Norwegian speakers and Norwegian adjectives, it is important to note that there is considerable overlap in defective adjectives between Norwegian and Swedish. If an adjective is defective in Swedish, it is likely to be defective in Norwegian, and vice versa. Thus two communities of speakers agree, which suggests that the gap is not accidental.

We would, as a consequence of the very small number of exceptions, assume that there is a similar process at work in both languages causing the morphological defectiveness, and that at least some amount of generalization in what we observe is appropriate. Ideally, we would want to arrive at a model that can account for morphological defectiveness in both Norwegian and Swedish, as the phenomenon is so similar in these languages.

During grammaticality judgments, subjects have time to consider many other factors, such as norms and what they think is expected of them in the experiment. Obviously, the problem would not exist if people did not report that they felt uncomfortable with using the missing adjective forms. The *apparent regularity* of defective adjectives, meaning that it is possible to predict which words will lack a neuter form, suggests that these paradigmatic gaps are productive (cf. Pettersson, 1990, inter alia). Any new adjective with the right combination of form and semantics would predict that people would hesitate to form its neuter gender form. This paper builds on the mechanism suggested in Johansson (2003), which we argue is maximally able to account for the defective paradigms as they occur in both Norwegian and Swedish. Our addition to the model is that we consider pseudowords as well as real words.

Previous theoretical work (Fanselow & Féry, 2002; Lowenadler, 2010; Raffelsiefen, 2004; Rice, 2007; Pettersson, 1990) suggests phonological, semantic, and frequency factors, and the topic has not been conclusively settled. The speakers of languages that contain paradigmatic gaps somehow have acquired knowledge that these forms are indeed not used and lack alternatives. This is an extreme case of the *learnability problem*: how can you stop the generalization of a rule if there is no negative evidence? The topic of paradigmatic gaps is therefore of interest to the question of how speakers acquire language (cf. Baker's paradox; e.g., some verbs do not use the dative shift). The interested reader should, for example, look at the connectionist debate that had at its core the nature of generalization and how generalization could be constrained. Connectionist models are currently returning in the form of Deep Learning that is now able to handle much larger datasets than were available in the 1990s, using computational power that was not available in that decade.

This study is primarily concerned with the processing status of the missing forms. The missing forms will typically not occur reliably in corpus material. The fact that they are missing may be inferred from their deviation from the statistical expectations of occurrence inferred from the observed frequency of the common gender, which is in line with the distributional suggestion of Yang (2015). There is a difference between not observing a phenomenon and the phenomenon not existing. If a model, such as a rule, predicts significantly more neuter forms than observed, then we clearly have a statistically interesting anomaly. By analogy, if we had a model that predicted that we would observe many unicorns in a context and we cannot find any unicorns, then that observation would be strong evidence against that model.

The null hypothesis is that the adjectives we are interested in behave like other adjectives. If our anomaly is shown to be productive and predictable, then we have a linguistically interesting phenomenon, because we would be able to make correct predictions about where the rule will fail. This is essentially the linguistic version of The Duck Principle presented by Hornstein and Nunes (2014): "If something walks, talks and defecates like a duck, the default position is that it is a duck: i.e. If constructions α and β have the same properties, the grammar should generate them in the same way." In our case, if an adjective looks like any other adjective that has a neuter form this adjective should also have a neuter form. We predict that there are other factors that prevent these neuter forms. If these neuter forms are prevented, we predict that our participants will not be able to activate the lemma form of the adjective in question, and thus that there will be no priming effects even though normal adjectives have reliable priming effects.

We predict that these neuter forms will not prime their common gender form. Regular adjectives will have a neuter form that reliably facilitates the recognition of the adjective. Note that the question is not if we can write a string with an ending, the question is if that form will be able to reliably activate the mental unit carrying the meaning (i.e. the lemma), which is thought to be responsible for facilitation effects of morphologically related words. If the forms are simply missing by accident, which we suggest they are not, then we would observe that speakers are able to generate them and recognize them, through applying the neuter-forming rule without hesitation, and in that case we would expect those neuter forms to facilitate the recognition of their adjective.

We suggest that experimental methods are capable of investigating linguistic intuitions through the phenomenon of priming, since differences in automatic flow of processing are shown in differences in reaction times and error rates, both of which are hard to control consciously by the subjects. In short, priming gives access to a measure of how words are related to each other in the mental lexicon. Since responses in a lexical decision task are hard to consciously manipulate, due to the priming stimuli being shown only for a fraction of a second, we argue that the measurement is a good indicator of automatic linguistic processing (cf. Levelt et al., 1999; Levelt, 2001). This is even clearer for the masked priming paradigm (cf. Forster & Davis, 1984b), where the subjects may not be aware that they saw a priming stimulus. Priming will be introduced in more detail in the methods section.

2. Experimental materials

The regular class consists of adjectives that regularly form neuters. Neuter gender in adjectives in Norwegian is only morphologically distinct in the indefinite singular, and gender distinctions are collapsed in all other inflectional categories (in standard Norwegian). The problematic class consists of adjectives with defective paradigms lacking a neuter indefinite singular form.

Norwegian adjective morphology is, for the most part, simple and predictable. In order to form a neuter form of an adjective, take the stem of the adjective and add the neuter morpheme *-t*. The process is only slightly complicated by predictable processes of assimilation and a concomitant change in vowel quantity. There are some exceptions, but those exceptions follow regular patterns. To qualify as a candidate of the defective paradigm, a typical single-syllable adjective must either have no rime in the coda, or a rime containing an alveolar stop, /t/ or /d/, as highlighted in examples (1)–(4).

(1) *En flat hund.*
a.M/F.SG.INDEF flat.M/F.SG.INDEF dog.M.SG.INDEF
'A flat dog.'

(2) *Et flatt menneske.*
a.N.SG.INDEF flat.N.SG.INDEF human.N.SG.INDEF
'A flat human.'

(3) *En lat hund.*
a.M/F.SG.INDEF lazy.M/F.SG.INDEF dog.M.SG.INDEF
'A lazy dog.'

(4) *Et *latt menneske.*
a.N.SG.INDEF lazy.N.SG.INDEF human.N.SG.INDEF
'A lazy human.'

A smaller group is the disyllabic adjectives with main stress on the second syllable. While it is not immediately clear why a neuter form (*flatt*) is allowed but not another (**latt*), the ill-formedness appears to be pseudo-regular in other ways. There is a strong tendency for these adjectives to share semantic features relating to internal states or non-observable properties (cf. Johansson, 2003; Johansson & Torkildsen, 2005; Pettersson, 1990; Lowenadler, 2010), such as mental states (e.g., laziness, happiness, and pride).

There is no repair for these defective paradigms. Speakers are forced to use a different adjective with roughly the same meaning, substitute the noun so that neuter agreement can be avoided, or in extreme cases use a form that does not sound correct to them. The missing neuter forms may thus occur as 'occasionalisms' in (written and) spoken language. The missing forms are not different words with a separate meaning, so it does not take much effort to discern the speaker's intention. Even so, that could be said as well about removing the distinction between long and short vowels for many words, as this too would result in identifiable words that may 'sound' incorrect. In a corpus search (cf. Knudsen & Fjeld, 2013), we found that the problematic neuter forms are virtually non-existent in writing, even for adjectives that are highly frequent in their common gender forms.

Our stimuli are grouped into three classes: regular adjectives, defective adjectives, and pseudo-words. The stimuli used in our experiment are coded: regular, problematic, and non-word. The class non-word are all *pseudo-words*, as non-word is a broader category. Pseudo-words are used as a stricter label for the non-word decision, as they follow the phonotactic rules of existing Norwegian words. Thus, it is a more realistic challenge for our subjects to separate words from pseudo-words. The lexical classes were balanced using 15 regular adjectives and 15 adjectives with defective paradigms, selected as pairs with similar expected relative frequencies of the common gender form. We constructed 30 pseudo-words to match this, in order for the task to be balanced between word and non-word decisions. The following short lists present the stimulus items. A table that includes the neuter forms and translations, as well as all the pseudo-adjectives with their neuter forms, is found in the appendix.

Regular: *hvit, vid, høy, flat, god, bred, rå, fet, død, blå, ny, fri, het, brå, bløt*
Defective: *lat, sta, gravid, glad, solid, distré, redd, sky, tru, slu, staut, vred, kry, kåt, ru.*

For the non-word stimuli used in this study, we constructed pseudo-words without semantic content. Owing to the productivity of the neuter-forming rule, we can further generate expected neuter forms of these pseudo-words to use as priming stimuli for our study. We matched the phonotactic patterns to be similar across our three classes (normal, problematic, and nonwords).

While there may be some disagreement among native speakers on which specific items should be included in the defective category, our list is based on items that show phonological and semantic regularities and whose neuter gender forms do not significantly occur in corpus data. There may also be potential dialectal variation that we have been unable to account for. We will handle this statistically by treating subjects and items as random effects (sources of variance).

Common gender forms of adjectives were used as targets for the lexical decisions. These are well-formed in all cases for both regular and problematic adjectives.

All of the lexical items selected for the study are familiar to most Norwegian speakers in their common gender forms. This ensures that participants are able to recognize them during the task. There are more adjectives that would fit the criteria, but some are less frequent and some are dated in modern Norwegian.

We will now continue to introduce the priming paradigm used in this study, followed by a discussion of the participants and the equipment. Next follows a presentation of the results. Our interpretations are discussed before the presentation of the final conclusions.

3. Method

We employ a straightforward lexical decision task (cf. Schvaneveldt & Meyer, 1973), where participants were asked to classify Norwegian adjectives and pseudo-words (pseudo-adjectives) as words or non-words. Neuter forms were used as priming stimuli for this decision. Hypothetical neuter forms could easily be generated for all the pseudo-words and defective adjectives, due to the predictable nature of neuter-morphology in Norwegian. All decisions are made for the common gender form, which are all unproblematic as words. We compare the effect of priming by measuring reaction times for both primed and unprimed, counterbalanced for presentation order. We are particularly concerned with the priming effect within each class, compared across classes. The priming effect is the effect on reaction time attributable to showing the (possibly hypothetical) neuter gender form for a fraction of a second before the decision. Statistical techniques will allow us to estimate if there is an observed difference between primed and unprimed events. If a difference is found we can determine if the difference is facilitation, i.e. a speedup of processing for the lexical decision, or possibly a slowdown of processing, possibly related to activating a decision for a non-word. The statistical null hypothesis is that there is no difference in priming of regular and defective adjectives, and they are thus affected equally by priming with their neuter form.

3.1 Design

Speakers are tasked with classifying stimuli as either a word or not a word. Pseudo-words were used for the non-word stimuli. We used masked priming (cf. Forster et al., 1987), as this make it more difficult to consciously experience the priming stimulus and use the stimulus to calculate the intended form. In masked priming, subjects are less likely to report seeing the priming stimulus, and the priming episode is done in a fraction of a second, leaving little room for problem-solving strategies. Neuter forms of adjectives are used as priming stimuli for their corresponding common gender forms, as the lexical status of the common gender form is certain, so there is no doubt about the correct response for each item. Similarly, for the pseudo-words, we use easily generated hypothetical neuter forms. This provides a consistent test condition within each stimulus category. Activation of the target item can follow from form similarity and meaning similarity through the activation of a common lemma form. Pseudo-words will not have lemma forms, so their activation, if any, will be mainly through form similarity.

A warm-up lexical decision task was presented to participants in order to familiarize them with the procedure. The warm-up task was integrated seamlessly into the rest of the experiment so that there was no pause between the warm-up phase and the actual experiment. By familiarizing subjects with the demands of the task, we should see a lower number of extreme responses that would have to be excluded in order to avoid biasing the model.

Words were displayed on a 4K UHD screen. Participants used a Cedrus response pad to perform the task, using a green key for real word decisions and a red key for non-word decisions. The experiment was conducted in a soundproof room, on a Mac Mini, using SuperLab v5.1.

All collected decisions were made within one second after seeing the target word. This was trained in the warm-up session, and it was generally not a problem for participants to make a decision within that time limit. We specifically checked for outliers for both subjects and items, and we only used the correct decisions from the subjects (i.e. word for regular and problematic, and non-word for the pseudo-word).

Two kinds of baselines

The priming effect is estimated using an unprimed baseline for all classes. It is very hard to find words that are unrelated for all factors of meaning and form. In this experiment, we are comparing the priming effect between regular and defect paradigms, and the only baseline that is certainly equally unrelated to both classes is showing no word at all. If we had shown different unrelated words the effects could have been due to the different sets of baseline words. Each participant made a decision on *both* primed and unprimed versions of all items during the course of the experiment. The order of presentation was randomized and *counterbalanced*, such that half of our participants performed the baseline task first and the other half performed it second. This means that participants were exposed to each individual target stimulus twice, but this training effect will cancel out, across all participants. The use of masked priming makes it very difficult for participants to recall when and if a prime word was presented. The other kind of baseline is to compare the priming effect within each stimulus class against the priming effect in other classes. Will the priming information influence decisions equally over all classes?

The primary measure is the reaction times recorded for each decision of our participants. If there is an effect from priming with neuter forms, we expect to observe this in faster reaction times for the primed section of the experiment. We used a two-tailed evaluation in the analysis. However, we predicted no priming for the problematic adjectives based on previous research on both Swedish and Norwegian in our lab (cf. Johansson, 2003; Johansson & Torkildsen, 2005).

For the priming stimuli, we used dedicated forward and backward masks, consisting of # marks. The masks were presented in the unprimed baseline experiment as well, even though there was no stimulus (i.e., a blank screen for 50ms) between them. The masks were presented for 100 milliseconds before *and* after the prime, and the priming stimulus was presented for 50 milliseconds. Once the target stimulus was displayed, participants had one second to make their decision. A failure to classify the stimulus in time was recorded as a non-response. The sequence repeated until all controlled stimuli had been presented. All stimuli started with a fixation point (*) that was presented for exactly one second, then the masked prime, and finally the target word until time-out or a decision was made. Reaction times were collected and labeled by subject, item, class, and condition.

3.2 Participants and procedure

We recruited 30 native speakers of Norwegian to participate in the experiment, all from the student population at the humanities building of the University of Bergen. As a group, they were typical of the student population, with an estimated age range of 20 to 35 years old. None of the participants were asked to identify themselves, and all participation was voluntary and unpaid.

The main criteria for inclusion were that subjects were native speakers of Norwegian (self-reported) and that they performed above two thirds correct on all the included test material. Responses faster than 350ms were excluded. The demands of the task make it unlikely that a decision can be faster, considering the time it takes to send a signal to the finger to press the key. The actual procedure involved a careful statistical analysis of the rate of correct answers, after pruning the fastest responses. Two subjects had a significantly higher error rate. One of them had fewer correct answers than non-correct (which included too slow responses and incorrect responses). One of the excluded subjects was also significantly slower in making the decisions. All other participants had above 70% correct responses. After the outlier analysis, 28 subjects remained.

Participants were briefed on how to perform the task and informed about consent and that they were free to leave at any time, for any reason. Information relating to consent was displayed at the beginning of the experiment, and participants were additionally informed verbally beforehand. Participants were debriefed at the end of the task and given a chance to ask questions related to the experiment.

Participants made their decisions using a response pad to classify lexical items that appeared on a computer screen. They were instructed to press a green key on the response pad to classify the stimulus item as a word, and a red key to classify the stimulus item as a non-word. The experiment was conducted in a soundproof recording studio. Participants generally took 10 minutes or less to complete the experiment.

4. Results

Reaction times from the experiment were analyzed using linear mixed effects models (cf. Baayen et al., 2008), implemented in R by the lmerTest package (Kuznetsova et al., 2016). The mixed effects used priming condition (primed or not primed) and stimulus class (regular, problematic, or pseudo-word) as fixed effects, with random effects for subjects and items. Since we considered that class and condition could affect participants differently, we added specific slopes for these factors. This is feasible, as each participant gives repeated measurement points for each level of class and condition. For items, we only considered that items could be reacted to differently. Each item belongs to one and only one class: regular (normal), problematic, or pseudo-word (non-word).

A careful statistical outlier analysis showed that one of the problematic adjectives (*ru* 'rough, coarse') was significantly more often misclassified as a non-word. Its hypothetical neuter form is *rutt*, which is Swedish for 'route' and similar to Norwegian *rutte*, meaning to waste or spend. Both would possibly prime for another lemma than the *ru* 'rough'. We excluded '*ru*' and rebalanced our dataset by similarly excluding the most erroneous item in the other classes (see appendix for the full data set). This avoids effects that are potentially driven by extreme items that are not representative of the class. After exclusion and rebalancing, we are left with 56 different items: 14 in each adjective class, and 28 pseudo-words.

After removing the no responses and responses faster than 350ms, we have 2641 remaining data points.

The model is defined as follows using R syntax:

RT ~ Class*Condition + (1 | Item) + (Class*Condition | Subject)

Running an ANOVA on this model, after the within-degrees of freedom has been estimated using Satterthwaite approximation, we find significant effects for both fixed variables, as well as their interaction (cf. Table 1). There are significant main effects of priming condition (**), and class (***), and priming is significantly different for at least one class (*), shown in Table 1.

Table 1. ANOVA table for the L1 model. (SS = summed squared deviance, df = degrees of freedom)

	Mean SS	Between df	Within df	*F*	Pr (>F)	
Class	320387	2	56.96	34.83	< 0.001	(***)
Condition	102652	1	26.97	11.16	< 0.01	(**)
Class by Condition	32066	2	33.46	3.49	< 0.05	(*)

We used a post-hoc GLHT analysis (cf. Bretz et al. 2010) of our model to locate the effects. Table 2 is an excerpt from a larger table testing all unique combinations (6*5/2 = 15) of class and priming against each other. The Holm method was used to compensate for multiple comparisons, which is reasonable especially since we are only interested in three particular contrasts of priming within classes. The significance values are more conservative than if we had tested only these *selected* three combinations.

Decisions for non-words are 35ms faster when primed by their hypothetical neuter. Regular adjectives have a strong priming effect of 47 ms. The most interesting result is that the problematic adjectives do not have a significant priming effect (16ms). The lack of priming confirms our prediction that the problematic neuter forms will not activate their lemmas, and thus they will not facilitate recognition of their adjective in any other way than by similarity in form. The small, and statistically insignificant, effect is readily explained by similarity in form.

Table 2. Pairwise comparisons (Primed – Unprimed)

Priming	Estimate	Std. Error	z value	Pr (>|z|)	
Nonword	34.54	11.32	3.05	0.025	(*)
Normal	47.24	12.33	3.83	0.002	(**)
Problematic	16.24	12.39	1.31	0.74	n.s.

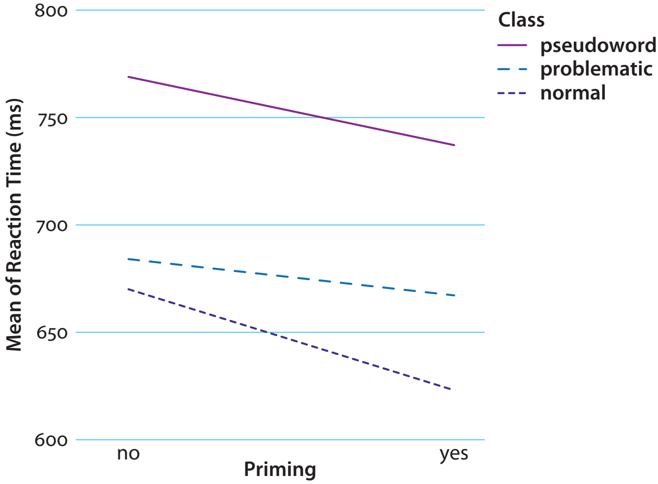

Figure 1. Interaction of reaction times

The slower reaction times for primed and unprimed problematic adjectives are not significant ($p > 0.7$) but somewhat slower reaction times are expected as these adjectives are not concrete and refer to mental properties that may be somewhat harder to process. However, this effect is not related to priming.

Our pseudo-words are generally slower than both groups of lexical items, which is expected due to the word superiority effect (Paap et al., 1982). Figure 1 summarizes the findings: a significant priming effect for non-words and for regular adjectives, and simultaneously no significant priming effect for the problematic adjectives, as predicted.

5. Discussion

Priming with the neuter form made lexical decisions significantly faster for regular adjectives. However, we did not see any facilitation effect for the problematic adjectives, as predicted if the neuter forms were not able to activate the adjective through lemma activation.

Looking closer at the pseudo-words, we observed a significant priming effect for correct lexical decisions rejecting their word status. That effect cannot be mediated by activation of an existing lemma for pseudo-words. Considering that form similarity may have significant effects, we note that similar priming effects for the problematic adjectives were absent.

The priming stimuli used in the experiment were both graphemically and phonologically similar enough that we expected some level of activation due to the form-priming effect (cf. Forster & Davis, 1984a). This may explain the slightly faster decisions for primed versus unprimed pseudo-words.

If form priming is present for the pseudo-words we used in the experiment, then we would assume this effect to be present in our problematic adjectives as well. The results found no significant priming effect for problematic adjectives, but such an effect was reliably detected for the ordinary adjectives.

It may be the case that the non-words prime a decision for non-words, that is, form similarity makes it possible to recognize the pseudo-word already when the prime stimulus is presented. The participants could recognize the priming stimulus as missing from the lexicon, and this would facilitate the decision for non-word.

In the problematic class, if our participants had been primed to make incorrect decisions for non-words, by priming with the problematic neuter forms, then we would expect that there could be more errors when primed. Since no responses could be used as a strategy and since the number of errors was low, we put erroneous decisions and no response into one category, i.e. non-correct.

Table 3. Error analysis. Non-correct in each category

	Normal	Problematic	Pseudo-word
Primed	6.1% (24)	14.3% (56)	21.8% (171)
Unprimed	3.3% (13)	11.0% (43)	24.0% (188)

When we analyze the non-correct responses in the data (after removing outliers), we notice that the primed responses are less correct than the unprimed for the two adjective classes and more correct for the pseudo-words (cf. Table 3). However, a statistical analysis of the table finds no significant association between Class and Priming for incorrect responses. This result can be explained by random variance, a trade-off between speed and accuracy, and the closeness of pseudo-adjectives to existing adjectives. Pseudo-words were simply hard to decide on correctly as they were so similar to real words, and therefore priming helped decisions more for non-words. One interpretation of the lack of a reliable association is that our subjects were performing well on the task, after removing outliers.

6. Conclusions

From the discussion, we conclude that our method has been successful in achieving what we set out to do, namely to investigate paradigmatic gaps through priming effects. The intuition of native speakers is that the neuter form does not exist, and our priming experiment confirmed that the problematic neuter forms did not significantly facilitate the adjective, which we interpret as their neuter forms failing to activate their lemma form.

This method is particularly useful when investigating linguistic features where speakers do not agree on the status of some phenomenon, or where their introspection would be highly variable, perhaps based on knowledge of normative rules acquired after childhood.

The lack of priming for defective adjectives in the experiment is note-worthy. We would expect form-similarity to cause at least some degree of facilitation, since the difference between common gender and neuter forms is the presence of a *-t* added at the end. There is of course a chance that form priming could be found as a weak effect if the study was much larger.

We observed facilitation for the decisions of pseudo-words as non-words, which used constructed neuter forms as the priming stimuli. Since pseudo-words do not have any semantic or associative content, the facilitation indicates form-similarity as the most likely explanation for pseudo-word priming in our experiment: It is easier to take the decision for a non-word if no adjectives were activated by the constructed pseudo-neuter and the rejection of word status can be achieved sooner when the (neuter form of the) pseudo-word was exposed already as a prime.

We indicated that the paradigmatic gap is productive. In future research, we will use novel pseudo-words that can be introduced with different meanings. The same pseudo-word may be introduced with meanings either satisfying the semantic criteria of being a hidden (mental) property, or being a concrete observable attribute. Our prediction is that those two pseudo-classes of the same novel word will have different priming effects for lexical decisions. Those subjects exposed to a non-concrete meaning will get no priming effect, and those exposed to a concrete meaning will have faster responses. In training, only the common gender form will be used for assigning meaning to the pseudo-words. This will be a version of the classic wug-test (cf. Berko, 1958), but we would like to investigate this through the lens of priming effects. We suspect this allows a better window into the processing that underlies the phenomenon and avoids, or lowers, the possibility that the subjects develop strategies for how they think we want them to perform.

Acknowledgement

We would like to thank our anonymous reviewers for several useful comments that resulted in an extensive revision and restructuring of the article that we think has made it clearer. We also thank our department colleague Tori Larsen, and professor emeritus Robert James Jarvella for valuable corrections and encouragement after reading the penultimate manuscript.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baerman, M., Corbett, G. G., & Brown, D. (Eds.). (2010). *Defective paradigms. Missing forms and what they tell us*. Oxford: Oxford University Press & British Academy. https://doi.org/10.5871/bacad/9780197264607.001.0001

Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry*, 10, 533–581.

Berko, J. (1958). The child's learning of English morphology. *Word*, 14(2-3):150–177. https://doi.org/10.1080/00437956.1958.11659661

Bretz, F., Hothorn, T., & Westfall, P. (2010). *Multiple comparisons using R. Boca Raton*, FL: CRC Press.

Fanselow, G., & Féry, C. (2002). Ineffability in grammar. In *Resolving conflicts in grammars: Optimality Theory in syntax, morphology, and phonology* (pp. 265–397). Hamburg: Buske.

Forster, K. I., & Davis, C. (1984a). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language*, 30, 1–25. https://doi.org/10.1016/0749-596X(91)90008-8

Forster, K. I., & Davis, C. (1984b). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 10(4), 680–698.

Forster, K. I., Davis, C., Schoknech, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation? *The Quarterly Journal of Experimental Psychology*, 39(2), 211–251. https://doi.org/10.1080/14640748708401785

Hornstein, N., & Nunes, J. (2014). Minimalism and control. In A. Carnie, D. Siddiqi, & Y. Sato (Eds.), *Routledge handbook of syntax* (pp. 239–263). London: Routledge.

Jenset, G. B., & Johansson, C. (2013). Lexical fillers influence the dative alternation: Estimating constructional saliency using web document frequencies. *Journal of Quantitative Linguistics*, 20(1), 13–44. https://doi.org/10.1080/09296174.2012.754597

Johansson, C. (1999). Learning what cannot be by failing expectations. *Nordic Journal of Linguistics*, 22, 61–76. https://doi.org/10.1080/03325869950137054

Johansson, C. (2003). How is grammatical gender processed? In T. Gaustad (Ed.), *Computational linguistics in the Netherlands 2002* (pp. 65–76). Amsterdam: Rodopi. https://doi.org/10.1163/9789004334441_007

Johansson, C. (2017). A word or two? In V. Rosén & K. De Smedt (Eds.), *The very model of a modern linguist — in honor of Helge Dyvik* (pp. 112–126). *Bergen Language and Linguistics Studies,8*(1). https://doi.org/10.15845/bells.v8i1

Johansson, C., & Torkildsen, J. C. (2005). Reaktionstider för genuskongruens: Hur processeras genus? [Reaction times for gender congruence: How is gender processed?]. In S. Lie, G. Nedrelid, & H. Omdal (Eds.), *MONS 10: Utvalde artiklar frå det tiande Møte om norsk språk i Kristiansand 2003* (pp. 157–167). Kristiansand: HøyskoleForlaget.

Knudsen, R. L., & Fjeld, R. V. (2013). A balanced; annotated national corpus for Norwegian Bokmål. In *Proceedings of the Workshop on Lexical Semantic Resources for NLP at NODALIDA* (NEALT Proceedings Series 19). Tartu: DSpace at Tartu National Library.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). *lmerTest: Tests in linear mixed effects models*. R package version 2.0-32.

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America*, 98(23), 13464–13471. https://doi.org/10.1073/pnas.231459498

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75. https://doi.org/10.1017/S0140525X99001776

Lowenadler, J. (2010). Relative acceptability of missing adjective forms in Swedish. In M. Baerman, G. G. Corbett, & D. Brown (Eds.), *Defective Paradigms. Missing forms and what they tell us* (pp. 69–83). Oxford: Oxford University Press & British Academy. https://doi.org/10.5871/bacad/9780197264607.003.0005

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word-superiority effect. *Psychological Review*, 89(5), 573–594. https://doi.org/10.1037/0033-295X.89.5.573

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: The MIT Press.

Pettersson, T. (1990). Varför barnet inte kan vara latt [Why the child cannot be lazy]. In E. Andersson & M. Sundman (Eds.), *Svenskans Beskrivning, 17*, 293–302. Åbo: Åbo Academy Press.

Raffelsiefen, R. (2004). Absolute ill-formedness and other morphophonological effects. *Phonology*, 21(1), 91–142. https://doi.org/10.1017/S0952675704000144

Rice, C. (2007). Gaps and repairs at the phonology-morphology interface. *Journal of Linguistics*, 43(1), 197–221. https://doi.org/10.1017/S0022226706004488

Schvaneveldt, R. W., & Meyer, D. E. (1973). Retrieval and comparison processes in semantic memory. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 395–409). New York NY: Academic Press.

Yang, C. (2015). Negative knowledge from positive evidence. *Language*, 91(4), 938–953. https://doi.org/10.1353/lan.2015.0054

Appendix

Stimuli items in the different categories. Excluded items in boldface. The alternation between neuter ending '-dt' or '-tt' is a spelling convention rather than a phonological realization, both mark [t:] after a short vowel.

Problematic	**neuter**	**translation**	**Regular**	**neuter**	**translation**
lat	*latt*	lazy	*hvit*	*hvitt*	white
sta	*statt*	stubborn	*vid*	*vidt*	wide
gravid	*gravidt*	pregnant	*høy*	*høyt*	high
glad	*gladt*	happy	*flat*	*flatt*	flat
solid	*solidt*	solid	*god*	*godt*	good
distré	*distrétt*	absent-minded	*bred*	*bredt*	broad
redd	*redt*	afraid	*rå*	*rått*	raw
sky	*skytt*	shy	*fet*	*fett*	fat

Problematic	neuter	translation	Regular	neuter	translation
tru	*trutt*	true/dependable	*død*	*dødt*	dead
slu	*slutt*	sly	*blå*	*blått*	blue
staut	*stautt*	stubborn	*ny*	*nytt*	new
vred	*vredt*	angry	*fri*	*fritt*	free
kry	*krytt*	healthy/proud	*het*	*hett*	hot
kåt	*kått*	horny	*brå*	*brått*	sudden
ru	***rutt***	unpolished	***bløt***	***bløtt***	soggy

Pseudo-word	pseudo-neuter	Pseudo-word	pseudo-neuter
balid	*balidt*	*krødd*	*krødt*
brad	*bradt*	*løyd*	*løydt*
dispré	*disprétt*	*mergid*	*mergidt*
fled	*fledt*	*prad*	*pradt*
frøyd	*frøydt*	*rod*	*rodt*
glo	*glott*	*ryd*	*rydt*
glu	*glutt*	*saut*	*sautt*
glydd	*glydt*	*skly*	*sklytt*
grod	*grodt*	*skrø*	*skrøtt*
hadd	*hadt*	*sky*	*skytt*
hød	*hødt*	*tad*	*tatt*
vrut	*vrutt*	*termid*	*termidt*
kau	*kautt*	*tju*	*tjutt*
klu	*klutt*	*trei*	*treit*
fro	***frott***	***hvet***	***hvett***

Determiner Phrases

A good-enough representation is not good enough

Loes Koring
Macquarie University

It has been proposed that language processing invokes extra-grammatical heuristics in addition to, or instead of the computational system (e.g., Townsend & Bever, 2001; Karimi & Ferreira, 2016). The outputs of these extra-grammatical heuristics are called "good-enough" representations. These representations lack (syntactic) detail and are incomplete (Karimi & Ferreira, 2016). This paper evaluates this claim by investigating one extra-grammatical processing heuristic in particular: the NV(N)-strategy. Two experiments prove that (i) interpretations that would result from application of the NV(N)-strategy are sometimes difficult to generate and (ii) listeners compute (syntactic) representations for sentences that are more detailed than the NV(N)-strategy would predict. This gives rise to the question whether "good-enough" representations are computed at all.

Keywords: good-enough representations, extra-grammatical processing heuristics, parsing, unaccusativity

1. Introduction

One of the most important goals of psycholinguistics is to spell out what exactly happens from the moment we encounter an auditory stimulus (a spoken sentence) to the point at which we form a semantic representation for that auditory stimulus. Crucially, our brains do not take the transient auditory stimulus at face value. The incoming auditory signal must be recast into a more permanent representation. To determine how this is accomplished, we need to know what resources our brains draw upon in this process (cf. Koornneef & Reuland, 2016). One resource that could potentially facilitate the composition of a semantic representation is the grammar we have internalized in the course of language acquisition. It has been proposed, however, that the human sentence processing system (the parser) assigns meaning representations using extra-grammatical heuristics that are of a

https://doi.org/10.1075/lald.64.07kor

different nature (e.g., Christianson et al., 2001; Townsend & Bever, 2001; Ferreira, 2003; Ferreira & Patson, 2007; Bever, 2008; Karimi & Ferreira, 2016).

More specifically, the proposal is that the parser engages extra-grammatical processing heuristics in constructing meaning representations that are not based on grammatical knowledge. Sometimes, the output of the two routes (grammar vs. heuristics) is the same. In these cases, sentences are easy to parse. In fact, Townsend and Bever (2001) use Miller & Isard's (1963) findings that grammatical sentences are easier to repeat than ungrammatical ones as support for such a 'dual route' model: "words in sentences sound clearer, because they have two mental resonances, not one" (p. 185). The two mental resonances are (i) a representation that is the output of the individual's grammar, and (ii) a representation that is the output of the application of extra-grammatical heuristics (a pseudosyntax in Townsend and Bever's terms).

Heuristics are defined as mental shortcuts. They are hypothesized to be cheap and efficient strategies which can be used to generate meaning representations. For instance, Karimi and Ferreira (2016) assert that "[h]euristic processing saves effortful processing by proceeding through "fast and frugal" heuristics rather than slow going and computationally costly algorithmic processing [...]." (p. 1014). Clearly, this claim rests on the assumption that invoking grammatical knowledge is not cost effective (see also Koornneef & Reuland, 2016). If heuristics are cheap and grammatical computations are effortful, then using heuristics instead of grammar has a definite advantage. This raises the question if, and when, the parser is compelled to resort to grammatical knowledge. Indeed, Karimi and Ferreira (2016), referring to Ferreira and others, postulate that "[...] algorithmic procedures for sentence processing are not only too costly but sometimes outright unnecessary [...]" (p. 1014). On those occasions that the parser's meaning representations are the output of applying extra-grammatical processing heuristics, the representations are hypothesized to be "incomplete" or "lacking in detail." They are dubbed "good-enough" representations. The hypothesis that meaning representations are often just "good-enough" (outputs of extra-grammatical heuristics) is based on the observation that "[...] there is evidence that under some circumstances comprehenders do **not** obtain the meaning consistent with a sentence's true content" (Ferreira, 2003, p. 166). That is, sometimes the meaning representations listeners generate for sentences are not compatible with the syntactic parse of the (full) sentence (Ferreira, 2003; Karimi & Ferreira, 2016).[1]

1. Apparently, a (mis)representation of a sentence counts as good enough for Karimi and Ferreira (2016), but, as Koornneef and Reuland (2016) point out, a meaning representation that "is not consistent with the sentence's true content" is of course not good enough as a meaning representation. This is, however, the meaning representation that participants sometimes end up with and the crucial question is how participants arrive at such misrepresentations (Koornneef & Reuland, 2016).

This paper evaluates these general claims about extra-grammatical processing heuristics by looking in detail at one heuristic, namely the NV(N)-strategy. The NV(N)-strategy dictates that a sequence of expressions that form the combination NV(N) is assigned a meaning representation: agent + predicate (+ patient) (see e.g., Townsend & Bever, 2001; Bever, 2008). To evaluate this proposal, I will investigate whether our meaning representations are in line with the predictions that follow from such a strategy. The findings of two experiments reveal that:

i. (Good-enough) representations that result from applying the NV(N)-strategy were in some cases *difficult* for the participants to compute.
ii. The meaning representations the participants do compute were more detailed than the application of the NV(N)-strategy would predict.

This leads to the conclusion that "good enough" meaning representations are just not good enough. Because the meaning representations that participants created during language processing were more articulated than would be expected on the heuristic account, we are invited to ask whether such heuristics are ever used in sentence processing.

2. Processing heuristics

In order to evaluate the proposed extra-grammatical, alternative, heuristics employed in language processing, we need to know how they work exactly. It is problematic then to learn that "the nature of the simple rules that guide heuristic processing is unclear" (Karimi & Ferreira 2016, p. 1019; see also Koornneef & Reuland, 2016). The one heuristic that is reasonably clear is the NV(N)-strategy. This strategy is "[…] the comprehender's strong tendency to assume that the subject of a sentence is also the agent of some action and the object of the same sentence is the patient or theme […]" (Ferreira, 2003, p. 169). Despite the fact that this definition includes the grammatical notions *subject* and *object*, it is purportedly an extra-grammatical strategy.

Townsend and Bever (2001) provide some details about how the NV(N)-strategy works. In their model, there are different phases in processing. At the first phase, the parser isolates the major phrases. Taking sentence (1) as an example, this means that the parser articulates the input into the N-*Mary*, the V-*kicked*, and the N-*Bill*. At the second phase, the parser assigns a canonical sentence structure to this sequence. In the canonical structure, the first N corresponds to the agent of the action, the V that follows corresponds to the action, and the N that follows next corresponds to the patient or theme of the action (perhaps this is achieved by fitting a template to a linear string). The output of this strategy is the good-enough representation, one that results without applying "costly syntactic algorithms". In this case, presumably, the meaning representation that can be attained using the extra-grammatical heuristic mirrors the representation that is the output of the computational system.[2]

(1) Mary kicked Bill.

Although this strategy may sound intuitive and efficient, there are questions as to how it works exactly (for further discussion, see Koornneef & Reuland (2016)). One question concerns how the strategy is accessed. That is, is there a list of heuristics available that the parser browses to find the right heuristic to apply to the input? How many heuristics are there, and how does the parser decide which heuristic is the right one to use? Is there any cost to searching for the right heuristic? Without further details about the inner-workings of heuristics, it is not possible to assess the claim that the use of heuristics is cheap, as compared to syntactic algorithms.

Consider next the passive sentence (2). It is easy to see that the NV(N)-strategy would lead the parser astray in this case. The parser would end up with a meaning representation in which Mary is the agent of the kicking-action and Bill the theme, in contrast to its (grammatical) meaning. Interestingly, native English-speaking children and adults sometimes make mistakes in interpreting passive sentences, at least in certain experimental tasks (see, e.g., Ferreira, 2003). In certain tasks, the first N *is* indeed interpreted as the agent and the second N as the theme of the event. This is not only observed when people confront reversible passives as in (2), but it is also observed with *irreversible* passives like *the cheese was eaten by the mouse* (Ferreira, 2003).[3]

(2) Mary was kicked by Bill.

Note, however, that, in order to arrive at the (active) misrepresentation, the listener has to (a) perform some non-trivial syntactic computation and (b) ignore some crucial information in the sentence (Koornneef & Reuland, 2016). That is, in order for the first N to receive the role of agent, it must be integrated as the external argument of the verb (and the second N as the internal argument). In order to achieve this, the listener has to ignore *was* as well as *by* and treat *kicked* as the active, in-

2. In Townsend & Bever's (2001) model, this output is then checked against an output generated by using syntactic algorithms. The output of applying heuristics serves as a constraint for the grammatically produced output.

3. For Townsend and Bever (2001) it is not necessarily the case that the output of the computational system route is not generated, but errors might also arise if the two routes output distinct meaning representations.

stead of the passive form of the verb (see (2′)).[4] This makes the NV(N)-strategy indistinguishable from a reduced processing account in which the parser has insufficient data points to come up with the correct parse (Koornneef & Reuland, 2016). If the parser does not access the expressions *was* and *by*, the only possible parse that is licensed by the computational system is one in which Mary is the agent and Bill the theme.

(2′) Mary kicked Bill.

In short, there are some crucial questions related to extra-grammatical processing heuristics that remain unanswered so far. This makes it impossible to evaluate whether heuristic processing is indeed cheaper than using syntactic algorithms. Furthermore, in the case of the passive, the parser has to ignore crucial information in the input in order to apply the NV(N)-strategy. As such, the NV(N)-strategy is indistinguishable from a reduced processing account (Koornneef & Reuland, 2016). This makes it impossible to figure out whether an incorrect meaning representation is the result of a *different* processing strategy or the workings of our computational system with insufficient data.

The next section examines another case in which syntax guides the parser towards a different meaning representation than extra-grammatical heuristics. As will become clear, the extra-grammatical heuristic gives rise to a meaning representation that is in fact *difficult* for participants to get.

3. Another case in which the NV(N)-strategy fails

A simple case in which the NV(N)-strategy would provide an incorrect meaning representation is the case of unaccusative verbs such as *fall* (cf. Townsend & Bever, 2001). Unaccusative verbs assign the role of theme to their single argument, which appears in subject position (e.g., (3a)). The NV(N)-strategy would, therefore, output an incorrect meaning representation for sentences with unaccusative verbs, as it would produce a representation in which the subject of the sentence is the agent.[5] That is, a meaning representation in which, in our example in (3a), the

4. In contrast to Ferreira (2003) and Karimi & Ferreira (2015), for Townsend and Bever, the parser does make use of this information initially and the parser first applies the schema for "simple predication" which would lead to a correct hypothesized meaning.

5. Townsend and Bever (2001) argue that you do not apply the heuristic in this case, as the information of the particular verb tells you that the thematic role that needs to be assigned is a theme role. As research has shown, however, thematic role assignment crucially involves syntactic structure: a role can only be assigned in a particular syntactic position (e.g., Koring et al.,

boy purposefully fell, which is not the first meaning representation that comes to mind. Such a representation, however, is imaginable and can be achieved in so-called stage contexts in (3b) for instance (Hackl, 1998). This particular reading, however, seems more difficult to get. This is in stark contrast to sentences with unergative verbs such as *jump* in (4) in which the subject bears the role of agent.

(3) a. The boy fell.
  b. The boy can fall really well on stage.

(4) The boy jumped.

Given that an agentive falling-event is a possible parse of sentence (3a) *and* it is compatible with the NV(N)-strategy, it seems that this reading should be easier to get than a reading in which the boy is the theme of the falling-event. In Experiment 1 we tested whether participants accept agentive readings of sentences as in (3a). That is, Experiment 1 seeks to find out how well the NV(N) strategy reflects listeners' understanding of sentences with unaccusative verbs.

3.1 Experiment 1

Twenty-six typically developing children and ten adults participated in the experiment. The children ranged in age from 4;0–5;7, with a mean age of 4;8. The adults ranged in age from 24 to 41 years old. The participants carried out a felicity judgment task in which they had to judge which one of two sentences sounded better. A sample pair of sentences is given in (5). In each pair, we contrasted an unaccusative and an unergative verb. Both sentences triggered an ability reading for modal *can* by adding the phrase 'really well' to the test sentences. This encouraged the participants to assign an agentive reading for the verb.

(5) a. The girl can hide really well.
  b. #The girl can disappear really well.

Sentence pairs were presented in written version to adults. The pairs of sentences were presented aurally by a hand puppet to the children. The children were told that the hand puppet was just starting to learn how to speak English and she found it difficult to judge what is silly to say and what is okay to say. The child's job then

2012) (see also Koornneef & Reuland, 2016). As such, the role of theme will not be assigned without the accompanying syntactic structure. Furthermore, this gives rise to the question when and on the basis of what you decide to use a heuristic. Also, it seems inconsistent with what is argued for earlier in the book: namely that you assign the role of agent to an N as soon as the N has been perceived (without any further information). At that point, the parser does not know yet that the verb is going to be unaccusative.

was to help the hand puppet by telling her which one of the two sentences was silly. We presented the participants with a total of eight test sentence pairs.

Because the felicity judgment task can be difficult for children, we included four control items, such as the pair in (6). Children were only included in the analysis if they managed to judge three out of four control pairs right. This led to the exclusion of nine children. In addition to the control and test items, we added four filler items to the experiment. The child was familiarised with the puppet and task before the experiment with three additional sentence pairs. The task was typically completed in about 10 minutes.

(6) a. The lolly tastes really good.
b. *The lolly eats really good.

The results (as graphically represented in Figure 1) showed that neither children nor adults had any difficulty distinguishing between the two sentences in the pair. That is, in a majority of the cases children (and adults) judged the agentively used unaccusative verb to be silly. A mixed-effects logistic regression model with participant, item and participant by condition random effects and condition as a fixed effect indeed revealed a significant effect of condition on response ($b = 2.89$, $SE = 0.57$, $p < .0001$).

This shows that, even for our child participants, there is something silly about using an unaccusative verb agentively. This 'silly' response results from the participants having to do some extra work to achieve an agentive reading at all, known as coercion. In particular, the participant has to ignore syntactic information from the particular verb and, instead, build an unergative structure in which the agent role can be assigned. This clash in instructions leads to a "silly" response. This is an unexpected result given the NV(N)-strategy, as the agentive reading would be the output of applying a heuristic instead of using our grammar. As such, an agentive reading should be easy to retrieve and, according to Karimi and Ferreira, often be the *only* reading we get. What we find instead, however, is that these representations are *difficult* to get, resulting in a 'silly' response.

One might argue that this is not a good test for use of processing heuristics, as the sentences we presented were too simple. That is, "[a] sentence's syntactic complexity seems to influence the extent to which misinterpretations occur." (Ferreira 2003: 167). Potentially then, computing the syntactic (and corresponding semantic) representation for our test sentences is so easy that the use of processing heuristics does not come into play. Such an argument, however, is flawed unless there is a clear definition as to what counts as syntactically complex and what does not. There is a way in which an unaccusative structure counts as syntactically complex. Namely, an unaccusative structure involves a dependency between the subject position and the internal argument position as represented in (7). That

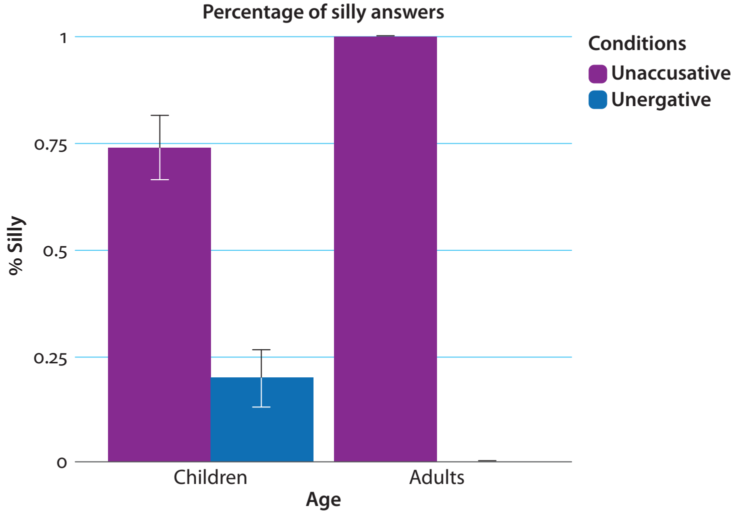

Figure 1. The percentage of *silly* responses participants gave for agentive unaccusative vs. agentive unergative sentences

is, even though the argument of an unaccusative verb appears in subject position, it displays (syntactic) behaviour that is typically associated with syntactic objects (e.g., Perlmutter, 1978; Burzio, 1986). Unaccusative structures, therefore, display the same type of (syntactic) complexity as passive sentences in which the verb's internal argument appears in subject position. This syntactic complexity gives rise to a distinct online processing pattern as compared to unergative verbs (e.g., Agnew et al., 2014, Bever & Sanz, 1997; Friedmann et al., 2008; Koring et al., 2012; Shetreet et al., 2010).

(7) $[_{TP}$ [The boy]$_i$ $[_{vP}$ fell t_i]]

Another question is whether we would expect children to use the NV(N)-strategy at all. According to Bever (1970) (and confirmed in Townsend & Bever, 2001), children acquire extra-grammatical processing strategies through a combination of experience, as well as universal constraints on the possibly available strategies. On the basis of children's performance on passives, Bever (1970) concludes that the NV(N)-strategy must be in place (and is in fact overused) at the age of four. This is based on the observation that children's performance on passives drops sharply at this age: "The most important feature of these results is the steady increase in performance until age 3.8 for girls and 4.0 for boys, when there is a sharp (temporary) drop in performance" (p. 31). The drop in children's performance is attributed to the NV(N)-strategy, which children are starting to apply, and are over-applying. Application of the NV(N)-strategy is not helpful in the case of the passive and, in fact, leads the children astray – resulting in the drop in performance.

The four-year-olds in the present study, then, should have adopted the NV(N)-strategy, and should be expected to be over applying it. If so, there should be nothing silly about any of the test sentences, which would make it difficult to distinguish between the two sentences in each pair. The finding that four-year-olds have very little trouble in judging which sentence is silly goes against the proposal that children at this age apply the NV(N)-strategy. In short, participants had a hard time retrieving meaning representations that correspond to the application of the extra-grammatical NV(N) processing heuristic.

The next section explores the meaning representations that emerge when participants are presented with ungrammatical sentences – the ultimate opportunity to apply extra-grammatical heuristics. The findings of this second experiment show that the resulting meaning representations are more detailed than would be predicted by extra-grammatical heuristics.

4. The meaning representations of ungrammatical sentences

The use of processing heuristics has been studied using grammatical sentences. But what happens when we present *ungrammatical* sentences to participants? Given that one might think that there is no straightforward way in which your grammar can provide you with a meaning representation, this would be the perfect opportunity to apply an extra-grammatical processing heuristic. In this section, we will therefore look at what meaning representations participants provide for ungrammatical sentences and, in particular, whether the output can be captured by the NV(N)-strategy.

In Experiment 2, participants were presented with ungrammatical sentences in which an intransitive verb appeared in a transitive structure, as in (8), inspired by Deal (2007). The task for participants was to provide a meaning representation for these sentences.

(8) a. Yesterday, I laughed a friend. (transitive unergative)
b. Yesterday, I slipped a friend. (transitive unaccusative)

Fourteen adult (24–30) control and 33 child participants (4;1–5;11) took part in this study. The (ungrammatical) sentences were presented to children by (silly) puppets on the computer. After the sentence had been produced, the experimenter would repeat the test sentence and ask the child: “what do you think he (the puppet) means?”. The child was then given two answer options (i.e. “do you think he means (a) or do you think he means (b)?). The answer options were provided in order to alleviate the (experimental) burden on the child participants and to make sure that also the less talkative children would provide us with sufficient data points. The two options given to the participants were (i) a causative structure (Made) and (ii) a structure that includes a prepositional phrase (PP) as in (8ab′). These answer options were based on the most frequent answers adult participants provided in a similar task as presented in Deal (2007). Participants were presented with a total of eight test items (four transitive unaccusatives and four transitive unergatives). These test items were combined with filler items of the type in (9) (novel verbs). In addition, there were items in which the puppets would just tell the child a silly joke. Adults received a written version of this task in which they read the test sentences and then chose among the two answer options.

(8) a′. <u>Made:</u> He made his friend laugh.
<u>PP:</u> He laughed at his friend.
b′. <u>Made:</u> He made his friend slip.
<u>PP:</u> He slipped over his friend.

(9) a. Yesterday, I lego-ed a castle.
b. Yesterday, I spaghetti-ed a friend.

The predictions of the NV(N)-strategy are straightforward: The first N is taken to be the (causal) agent of the action and the second N to be the (affected) patient. The strategy, however, does not tell you anything about how to best represent this in a structure (Made or PP). Your grammar, on the other hand, *does* give you a handle on how to best represent the corresponding meaning representation.

In particular, there is a more direct route from a transitive unaccusative to a causative representation than from a transitive unergative to a causative representation. In order to understand the difference, we need to go into the syntactic details of unaccusative, unergative and causative representations. The crucial difference between unaccusative (e.g., slip) and unergative (e.g., laugh) verbs is that unaccusative verbs assign their theme role to an *internal* argument, whereas unergative verbs assign their agent role to an *external* argument. A further relevant ingredient is a particular property of English such that a CAUSE projection only occurs with argument-complete VPs (a projection of a verb with an internal argument), but not bare verbs (Deal, 2007). Given this, a transitive unaccusative lends itself more easily for causativisation: We already have an internal argument position in which we can assign thematic role (i) (a theme role) and we assign a thematic (causer) role to the second argument in the external argument position. Alternatively, under some approaches, an unaccusative verb is in fact a *derived* entry that stems from a transitive entry with both a theme and a causer role (providing a direct route for the interpretation of transitive unaccusatives as causatives) (Reinhart, 2000/2016, 2002). The derived meaning representation is then most easily expressed as something like *He made his friend slip*.

In order to arrive at such a representation for transitive unergatives, on the other hand, we must *both* create an internal argument position (something that the unergative verb does not instruct the parser to do) and then, in an additional step, create an external argument position in which the role of causer can be assigned. This requires an extra step in retrieving such a representation, as compared to transitive unaccusatives. A causative meaning is therefore less straightforward for transitive unergatives. Alternatively, the parser might find an interpretation in which the subject is still the external (agent) argument of the verb and save the internal argument with a prepositional phrase (in which it gets assigned the thematic role of goal). Such a representation would be more faithful to the verb's argument structure. The prediction that follows from applying the rules of our grammar is therefore that the distribution of Made vs. PP responses will be affected by verb type. Namely, we will find proportionally more Made responses as compared to PP responses for transitive unaccusatives, but more PP than Made responses for transitive unergatives. No such prediction follows from applying the NV(N)-strategy.

The results from Experiment 2 are graphically represented in Figures 2 (children) and 3 (adults). For analysis of the child data, three children were excluded, as they consistently chose the second answer option offered as their answer, which can be taken as a lack of understanding the task. Figure 2 reveals that the type of verb included in the transitive sentence resulted in a different response distribution. More specifically, unaccusative verbs evoked more Made answers than PP answers, whereas unergative verbs evoked more PP answers than Made answers. Analyzing the data (excluding 'other' answers) using a multilevel regression analysis displayed a significant effect of condition on answer type ($b = -1.28$ (0.66), $p < .05$). Our final model included condition and age (four vs. five-year-olds) as fixed effects, but no condition by age interaction effect, given the absence of a significant interaction effect ($\chi^2(1) = 0.36$, $p = .55$). It included item and participant random effects.[6] There was no significant age effect ($b = -0.45$ (0.42), $p = .29$).

The result that verb type has an effect on the distribution of responses shows that our meaning representations are rich and detailed representations. Particularly, they are much more detailed than what the NV(N)-strategy can output.

6. I originally included participant by condition random effects as well, but this model was not significantly better than a model with participant and item effects only (χ^2 (2) = 1.91, $p = .39$).

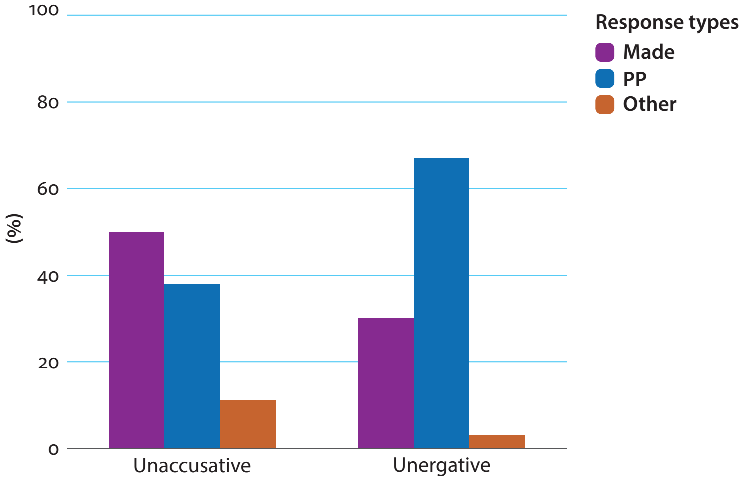

Figure 2. Percentage of answer types (Made, PP, other) per verb type (unaccusative vs. unergative) for child participants. Note that the category 'other' includes partial answers such as *his friend slipped* that could not be coded as either Causative or PP

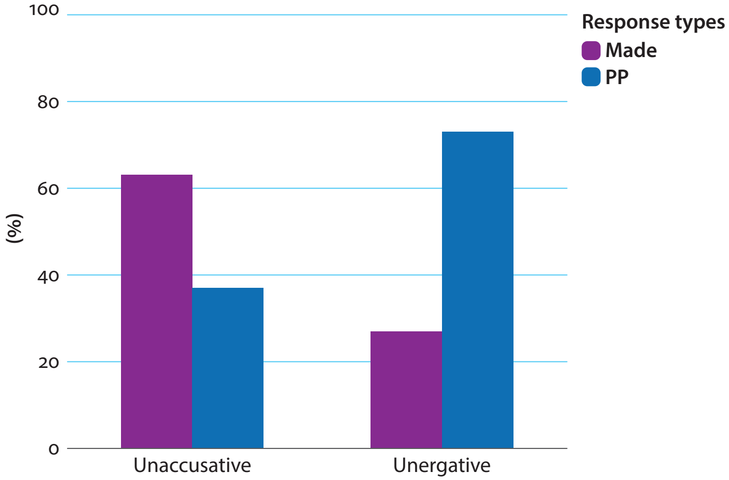

Figure 3. Percentage of answer types (Made, PP, other) per verb type (unaccusative vs. unergative) for adult participants

5. Discussion

The particular structure this study investigated to shed light on the use of extra-grammatical processing heuristics is the unaccusative structure. The choice for this particular structure is that it parallels *passive* structures in crucial aspects.

Participants' behaviour on passive structures has been used in support of extra-grammatical processing heuristics (e.g., Bever 1970, 2008; Ferreira 2003; Karimi & Ferreira 2016). The findings from experiments with both children and adults showed that passive sentences (as well as object-clefts) are harder to understand than active sentences (participants made more mistakes) and lead to longer processing times. These findings are then ascribed to application of the NV(N) strategy, which outputs the wrong results.

Passives and unaccusatives share particular structural characteristics. In both passive and unaccusative structures, there is a dependency between the subject of the sentence and the internal argument position (e.g. Burzio, 1986; Perlmutter, 1978). This is represented for unaccusative structures in (7), repeated here in (10a) and for passive sentences, as in (11b), in (12b). In contrast, in active sentences such as (11a) and unergative sentences, the argument in subject position is linked to the external argument position (see (10b) and (12a)).

(10) a. $[_{TP}$ [The boy]$_i$ $[_{vP}$ fell t_i]]
b. $[_{TP}$ [The boy] $[_{vP}$ t_i jumped]]

(11) a. The boy kicked the acrobat.
b. The acrobat was kicked (by the boy).

(12) a. $[_{TP}$ [The boy]$_i$ $[_{vP}$ t_i kicked the acrobat]]
b. $[_{TP}$ [The acrobat]$_i$ was $[_{vP}$ kicked t_i]]

The similarity between unaccusatives and passives on the one hand, and unergatives and actives on the other hand, is further illustrated by auxiliary selection in a language like Dutch (e.g., Hoekstra, 1984; Reinhart, 2002). Specifically, unaccusative verbs select the auxiliary *be* (13b), whereas unergative verbs select auxiliary *have* (13a). Like unaccusative verbs, passive verbs select auxiliary *be* (14b), whereas active verbs select auxiliary *have* (14a).

(13) a. De jongen heeft gerend
The boy has run
b. De jongen is gevallen
The boy is fallen

(14) a. De jongen heeft de acrobaat geschopt
The boy has kicked the acrobat
b. De acrobaat is geschopt (door de jongen)
The acrobat is kicked (by the boy)

The structural difference between unaccusative and unergative structures has repercussions in particular syntactic environments. That is, the subject of unaccusative verbs behaves syntactically like an internal argument, as such allowing ne-cliticisation in Italian (Burzio, 1986), possessive datives in Hebrew (Friedmann, 2007) and floating numeral quantifiers in Japanese (Miyagawa, 1989) for instance. The subject of unergative verbs is an external argument and, as such, this structure allows impersonal passives in Dutch (Perlmutter 1978) and pseudo-passives (Perlmutter and Postal 1984) in English for instance.

The similarity in syntactic "complexity" between unaccusatives and passives gives rise to the prediction that unaccusatives are equally difficult and slow for participants to process. That is, also in the case of unaccusative verbs, the NV(N) strategy would output an incorrect meaning representation (different from the output of the computational system). The findings of Experiment 1, however, showed that participants had no difficulty assigning the correct (syntactically computed) meaning representation, whereas the representation that would follow from the NV(N) strategy was difficult for participants to arrive at. That is, we did not find the corresponding mistakes that have been found in passive sentences in unaccusative sentences.

In fact, one might claim that computing the correct meaning representation for unaccusative sentences would lead to even more incorrect applications of the NV(N) strategy, given that there are no surface cues in unaccusative sentences that tell the parser that the underlying syntactic structure is unaccusative and not unergative (e.g., van Hout, 2004). In a passive structure, the crucial signals for the passive are the auxiliary (*was*) as well as the *by*-phrase. There are no such cues in unaccusative sentences. The absence in unaccusative sentences of the type of mistakes participants make with passive sentences, then, seems to support the hypothesis that the mistakes in passives are the results of reduced processing instead of extra-grammatical processing.

Finally, the present paper explored sentence processing with English-speaking participants, but the same predictions apply to other languages. A relevant source of data to this end is data on children's meaning representations for unaccusative vs. theme unergative (e.g. *stink*) type verbs. Sentences with unaccusative and theme unergative verbs look the same on the surface. Moreover, both verb types assign the same thematic role to their argument, yet their underlying syntactic representation is crucially different (Reinhart, 2000/2016). Results from an experiment showed that 4-year-old Dutch children distinguish between these verb types, suggesting that their meaning representations are, again, more detailed than the NV(N) strategy would output.

6. Conclusion

This paper evaluated the proposal that language processing sometimes makes use of extra-grammatical processing heuristics instead of (or in addition to) grammatical knowledge. These heuristics are claimed to be cheaper and more efficient than the outputs provided by the computational system for language. I pointed out that it is unclear whether extra-grammatical heuristics are indeed cheaper. Furthermore, I showed that one particular heuristic, the NV(N)-strategy, provides meaning representations for sentences that are difficult to arrive at. Finally, I showed that our meaning representations are detailed representations. Much more detailed than the NV(N)-strategy would predict. I therefore submit that there is no reason to compute any extra-grammatical (meaning) representations for sentences. People do not use good-enough representations; the meaning representations that people assign take full advantage of their internalized grammatical knowledge.

References

Agnew, Z. K., van de Koot, H., McGettigan, C., & Scott, S. K. (2014). Do sentences with unaccusative verbs involve syntactic movement? Evidence from neuroimaging. *Language, Cognition and Neuroscience*, 29(9), 1035–1045. https://doi.org/10.1080/23273798.2014.887125

Bever, T. G. (1970). The cognitive basis for linguistic structures. In R. Hayes (ed.) *Cognition and the development of language development* (pp. 279–362). New York, NY: Wiley & Sons.

Bever, T. G., & Sanz, M. (1997). Empty categories access their antecedents during comprehension: Unaccusatives in Spanish. *Linguistic Inquiry*, 28, 69–91.

Burzio, L. (1986). *Italian syntax: A government-binding approach*. Dordrecht: Reidel. https://doi.org/10.1007/978-94-009-4522-7

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42, 368–407. https://doi.org/10.1006/cogp.2001.0752

Deal, A. R. (2007). The asymmetry of argument structure: Evidence from coercion. *Semantics and Processing* (UMOP 37).

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47, 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

Ferreira, F., & Patson, N. D. (2007). The "Good Enough" approach to language comprehension. *Language and Linguistic Compass*, 1, 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Friedmann, N. (2007). Young children and A-chains: The acquisition of Hebrew Unaccusatives. *Language Acquisition*, 14, 377–422. https://doi.org/10.1080/10489220701600523

Friedmann, N., Taranto, G., Shapiro, L. P., & Swinney, D. (2008). The leaf fell (the leaf): The online processing of unaccusatives. *Linguistic Inquiry*, 39, 355–377. https://doi.org/10.1162/ling.2008.39.3.355

Hackl, M. (1998). *On the semantics of ability attributions* (Unpublished manuscript). MIT.

Hoekstra, T. (1984). *Transitivity: Grammatical relations in government- binding theory* (Unpublished dissertation). Rijksuniversiteit Leiden.
van Hout, A. (2004). Unaccusativity as telicity checking. In A. Alexiadou, E. Anagnostopoulou & M. Everaert (Eds.), *The unaccusativity puzzle: Explorations of the syntax-lexicon interface* (pp. 60–83). Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199257652.003.0003
Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal Experimental Psychology*, 69(5). https://doi.org/10.1080/17470218.2015.1053951
Koornneef, A., & Reuland, E. (2016). On the shallow processing (dis)advantage: Grammar and economy. *Frontiers in Psychology*, 7(82). https://doi.org/10.3389/fpsyg.2016.00082
Koring, L., Mak, P., & Reuland, E. (2012). The time course of argument reactivation revealed: Using the visual world paradigm. *Cognition*, 123, 361–379. https://doi.org/10.1016/j.cognition.2012.02.011
Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 2(3), 217–228. https://doi.org/10.1016/S0022-5371(63)80087-0
Miyagawa, S. (1989). Light verbs and the ergative hypothesis. *Linguistic Inquiry*, 20(4), 659–668.
Perlmutter, D. M. (1978). Impersonal passives and the unaccusative hypothesis. In J. J. Jaeger, A. C. Woodbury, F. Ackerman, C. Chiarello, O. D. Gensler, J. Kingston, E. E. Sweetser, H. Thompson, & K. W. Whistler (Eds.), *Proceedings of the fourth annual meeting of the Berkeley Linguistics Society* (pp. 157–189). Berkeley CA: BLS.
Perlmutter, D., & Postal, P. (1984). The 1-advancement exclusiveness law. *Studies in Relational Grammar*, 2, 81–125.
Reinhart, T. (2000). *The theta system: Syntactic realization of verbal concepts. OTS Working Papers in Linguistics*. Utrecht: University of Utrecht.
Reinhart, T. (2002). The theta system: An overview. *Theoretical Linguistics*, 28, 229–290.
Reinhart, T. (2016). The Theta system: Syntactic realization of verbal concepts. In M. Everaert, M. Marelj, E. Reuland, & T. Siloni (Eds.), *Concepts, syntax, and their interface: The theta system*. Cambridge, MA: The MIT Press.
Shetreet, E., & Friedmann, N. (2012). Stretched, jumped, and fell: An fMRI investigation of reflexive verbs and other intransitives. *NeuroImage*, 60(3), 1800–1806. https://doi.org/10.1016/j.neuroimage.2012.01.081
Shetreet, E., Friedmann, N., & Hadar, U. (2010). The neural correlates of linguistic distinctions: Unaccusative and unergative verbs. *Journal of Cognitive Neuroscience*, 22, 2306–2315. https://doi.org/10.1162/jocn.2009.21371
Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/6184.001.0001

Processing of pronoun gender by Dutch-Russian simultaneous bilinguals

Evidence from eye-tracking

Elena Tribushinina[1], Julia Lomako[2], Natalia Gagarina[2], Ekaterina Abrosova[3] and Pim Mak[1]
[1]University of Utrecht / [2]Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS) / [3]Herzen State Pedagogical University of Russia

This paper investigates the processing of pronoun gender by bilingual children. Prior research shows that Dutch–Russian bilinguals below age 7 often make gender agreement errors in Russian anaphoric pronouns, whereas monolingual children are target-like by age 4. This paper aims to establish whether the frequent production errors in the speech of bilinguals are due to lacking knowledge of grammatical gender or due to incidental performance breakdowns. The results of an eye-tracking experiment demonstrate that 5–6-year-old Dutch-Russian bilinguals are sensitive to gender cues, but are slower than Russian monolingual peers and adults in pronoun resolution. The findings support the view that bilinguals possess abstract grammatical representations, but are less efficient in gender production and processing due to competition cost in bilingualism.

Keywords: bilingualism, comprehension-production asymmetries, grammatical gender, anaphorical pronouns, online processing, eye-tracking

1. Introduction

Bilinguals usually hear and speak each of their languages less often than monolinguals. Reduced exposure to one or both languages may result in smaller vocabularies and/or delays in the acquisition of grammar. A grammatical phenomenon that appears to be particularly sensitive to the effects of reduced input is grammatical gender. Research has repeatedly shown that both L2 children (Blom, Polišenska, & Weerman, 2008; Blom & Vasić, 2011; Brouwer, Cornips, & Hulk, 2008; Hulk & Cornips, 2006; Keij et al., 2012; Orgassa & Weerman, 2008; Unsworth et al., 2014) and simultaneous bilinguals (Kupisch, Müller, & Cantone 2002; Rodina &

https://doi.org/10.1075/lald.64.08tri

Westergaard, 2017; Unsworth, 2013) make more gender agreement errors than their monolingual peers. Gender is also one of the domains where bilingual language development may even stagnate at a non-target-like level of performance (Hulk & Cornips, 2006; Franceschina, 2005; Unsworth, 2008), especially when the target gender system is opaque, as in Welsh, Irish and Dutch (Fhlannchadha & Hickey, 2017; Thomas & Gathercole, 2007; Unsworth et al., 2014). For example, Unsworth (2008) has found that some English-speaking learners of L2 Dutch persistently make gender errors by using the wrong determiner, even after significant exposure to the L2. For simultaneous bilinguals, Fhlannchadha and Hickey (2017) report that Irish-English bilingual children do not receive enough exposure to Irish to acquire the complex and opaque gender system and still make errors at age 13, especially with inanimate nouns.

How should frequent errors in the production of gender agreement be interpreted? One possibility is that bilinguals lack relevant grammatical representations or have a different representation of grammatical gender in their interlanguage systems (i.e., deviant grammars). In this case, we would expect a poor performance not only in production, but also in comprehension or online processing of grammatical gender. In line with this explanation, Keij and colleagues (2012) have found that child L2 learners of Dutch were outperformed by their monolingual peers in both a production task and a knowledge task (a forced-choice test of determiner-noun agreement). In a similar vein, Brouwer et al. (2008) report that 11 to 13-year-old bilingual children were outperformed by monolingual Dutch children in a grammaticality judgment task targeting gender agreement in the attributive domain. The participants were presented with either correct (e.g. *het schaap* 'the-NEUT sheep') or incorrect (e.g. *de schaap* 'the-COM sheep') determiner-noun combinations. Monolinguals were more accurate in identifying ungrammatical items and they also responded faster than bilinguals.

Another possible explanation of the frequent production errors is competition cost in bilingualism. Bilinguals have to deal with two competing language systems and suppress the language that is not currently in use (Bialystok, 2007; Paradis, 2010; Sorace & Serratrice, 2009). This additional cognitive load may result in less accurate language production than in monolinguals. A specific kind of production errors are transfer errors, i.e. errors made as a result of cross-linguistic influence from the other language of a bilingual. For instance, Nicoladis (2006) reports that English-French bilinguals overuse prenominal adjectives in French and postnominal adjectives in English. Such errors are consistent with the typological differences between English and French regarding adjective placement, and can therefore be seen as transfer errors. However, Foursha-Stevenson and Nicoladis (2011) have demonstrated that bilinguals are more sensitive to violations in adjective–noun order compared to monolinguals: In a grammaticality judgment task bilingual 5-year-olds performed significantly above chance in the recognition of ungrammatical adjective–noun orderings (57% correct), whereas their monolingual English-speaking peers performed at chance level (48% correct). These findings suggest that errors in adjective-noun ordering cannot be taken as evidence of lacking grammatical knowledge, but rather as an indication that bilingual children sometimes fail to inhibit the language that is not currently being used. In the same vein, Nicoladis (2006, p. 26) suggests that transfer errors are an "epiphenomenon of speech production" in the sense that bilinguals do have the appropriate knowledge of grammatical phenomena, but sometimes fail to inhibit the other language when they speak.

Similar production–comprehension asymmetries in bilingual children have been reported for subject–verb agreement in French (Ågren & Van de Weijer, 2013), tense marking in English (Chondrogianni & Marinis, 2012), discourse connectives in Russian (Mak et al., 2017; Tribushinina et al., 2017), pronominal anaphora in Russian (Gagarina & Sauermann, in press), definite articles in English and Dutch (Chondrogianni, Vasić, Marinis, & Blom, 2015), definite articles and clitic pronouns in Greek (Chondrogianni, Marinis, Edwards, & Blom, 2015). Unsworth (2013) arrives at a similar conclusion with regard to grammatical gender. In her study, bilinguals performed better on a grammaticality judgement task than on production of gender agreement in Dutch, which leads to the conclusion that errors might be "a production-specific performance problem rather than a failure to acquire those grammatical features and rules and/or to specify certain nouns with the target gender feature" (p. 105). Blom and Vasić (2011) draw a similar conclusion based on the results of self-paced reading task, where L2 children slowed down upon hearing incorrect (neuter) determiners, which can be considered as evidence of gender sensitivity despite variable production of determiner–noun agreement. Overall, online measures appear to be more sensitive to what bilinguals actually know than offline grammaticality judgment tasks (Zufferey et al., 2015). For example, based on the data from the grammaticality judgment task in Brouwer et al. (2008) (accuracy rates around 40%) we would have to conclude that bilinguals perform at chance and have not acquired the system of grammatical gender at all. However, the same participants had longer response times on ungrammatical trials, which suggests that they *are,* in fact, aware of the gender distinctions and therefore respond to ungrammaticality.

2. The present study

This chapter aims to contribute novel insights into the production–processing asymmetries in the gender domain by extending this line of research to the gender of anaphoric pronouns. As demonstrated by the literature review above, the bulk of research on the acquisition of grammatical gender by bilinguals has only addressed agreement in the attributive (and less often verbal) domain. Gender agreement between a pronoun and its antecedent is a largely under-researched area in the literature on bilingualism. However, there is recent evidence that the acquisition of pronominal gender may be particularly demanding for children with reduced exposure to one of the languages (usually a minority language that is not maintained in the country of residence). Tribushinina and Mak (2016) report that Dutch–Russian bilinguals growing up in the Netherlands (whose input in Russian is usually limited to their mothers' speech and input provided by Russian weekend schools) start producing Russian pronominal gender correctly only by age 7, whereas Russian monolinguals (including children with developmental language disorders) virtually make no errors from age 4 onwards. The present paper targets online processing of pronoun gender by Dutch–Russian bilinguals, at the age when they perform around chance level in their production of Russian pronoun gender (age 5–6).

If errors in the speech of bilingual children are an epiphenomenon of speech production, rather than an indicator of incomplete acquisition or deviant representation, we should find sensitivity to gender cues in a receptive task. A grammaticality judgment task does not seem suitable, since it tends to underestimate the ability of bilingual children to process gender cues (cf. Brouwer et al., 2008; Keij et al., 2012). A self-paced reading task (cf. Blom & Vasić, 2011) does not seem appropriate either, because sensitivity to gender cues has to be assessed in children who have not yet started literacy instruction. A method that appears to be particularly suitable in this case is eye-tracking by means of the Visual World Paradigm. It is well-established that adults search for antecedents based on number and gender cues as they hear a pronoun (cf. Sanford & Filik, 2007). Adults can also predict a referent based on a gender-marked determiner (Brouwer, Sprenger, & Unsworth, 2017; Huettig & Brouwer, 2015). Likewise, Lew-Williams and Fernald (2007) used eye-tracking to investigate the ability of Spanish-speaking toddlers to use gender cues in the determiner and found that 3-year-olds started looking at the target picture faster if the gender-marked determiner was informative about the upcoming referent. Specifically for pronoun gender, Arnold, Brown-Schmidt, and Trueswell (2007) have demonstrated that by age 4 English-speaking children use information about biological gender in online processing of pronouns and that children are as fast as adults in identifying the correct referent based on pronoun gender. To the best of our knowledge, this method has not yet been used to study gender processing in bilingual populations and processing of *grammatical* gender in the pronominal domain. The present study will fill this gap.

In the experiment reported below, the participants see two pictures and hear a sentence where both referents are introduced. The second sentence starts with a (masculine or feminine) pronoun that is either ambiguous (same-gender condition) or informative about the antecedent (different-gender condition). Since 5-year-old monolingual children are target-like in the production of pronominal gender, we expect that they will also use gender cues in the processing of anaphoric pronouns in the informative condition and will look at the target picture upon hearing the pronoun. For their bilingual peers, two scenarios are possible. If their chance performance in production is due to incomplete acquisition of grammatical gender (representation problem), they will not be sensitive to gender cues in the processing of pronouns, which means that their performance in the different-gender condition will be the same as in the same-gender condition where the pronoun is ambiguous and may refer to either of the referents. However, if the frequent errors attested in prior research are merely performance breakdowns (or in Nicoladis' terms an "epiphenomenon of speech production"), bilingual children should increase the proportion of their looks to the target picture upon hearing the informative pronoun.

This chapter is structured as follows. Section 2 briefly discusses the properties and learnability of Russian grammatical gender. Section 3 describes the methodology of the present study. Results are reported in Section 4 and discussed in Section 5.

3. Grammatical gender in Russian

Russian has a three-way gender system of masculine, feminine and neuter. The masculine form is considered the default (Corbett & Fraser, 2000). Nouns agree in gender with adjectives, past-tense verb forms, demonstrative pronouns and possessive pronouns. Anaphoric pronouns also agree with the grammatical gender of their antecedent: Masculine nouns (e.g. *kot* 'cat') require the pronoun *on* 'he', feminine nouns (e.g. *sobaka* 'dog') agree with the pronoun *ona* 'she' and neuter nouns (e.g. *okno* 'window') with *ono* 'it'.

In most cases the gender can be easily predicted from the morphophonological form of the noun. Masculine nouns usually end in a consonant, most feminine nouns end in *–a* and neuter nouns end in *–o*/*–e* (see Table 1). The system is therefore mainly transparent; both gender assignment and gender agreement are fairly unproblematic for children and are usually acquired by age 4 (Gvozdev, 1961; Rodina, 2008).

However, there are a few exceptional categories that are less transparent. For example, about one fifth of all feminine nouns (see declension III in Table 1) end in a palatalized consonant (e.g. *len'* 'laziness') and are ambiguous between feminine and masculine, because there are also masculine nouns ending in a palatalized

Table 1. Declensional classes of Russian nouns

	I (M)			II (F)	III (F)	IV (N)
	stol 'table' (inanimate)	*kot* 'cat' (animate)	*kon'* 'horse'	*korova* 'cow'	*mol'* 'moth'	*okno* 'window'
NOM	stol-ø	kot-ø	kon'-ø	korov-a	mol'-ø	okn-o
GEN	stol-a	kot-a	kon'-a	korov-y	mol'-i	okn-a
DAT	stol-u	kot-u	kon'-u	korov-e	mol'-i	okn-u
ACC	stol-ø	kot-a	kon'-a	korov-u	mol'-ø	okn-o
INS	stol-om	kot-om	kon'-om	korov-oj	mol'-ju	okn-om
LOC	stol-e	kot-e	kon'-e	korov-e	mol'-i	okn-e

consonant (e.g. *ogon'* 'fire') (Honselaar, 2014). In this case, children have to rely on syntactic agreement markers and noun endings in oblique cases. Second, there are nouns with a mismatch between semantics (biological gender) and morphology. For example, the morphologically feminine noun *papa* 'daddy' requires masculine forms of modifiers, verbs and pronouns; so grammatical agreement should in such cases be over-ridden by semantic agreement. Finally, stem-stressed neuter nouns (e.g. *solnc-{e}*[ə] 'sun) are barely distinguishable from feminine nouns (e.g. *mašin-{a}*[ə] 'car'), because unstressed endings are phonologically reduced in Russian. In this case, learners have to rely on end-stressed modifiers and verbs, and also on noun endings in oblique cases. Such exceptional cases take longer to acquire, monolingual children aged 5–6 still make errors with exceptional cases of syntactic agreement, over-relying on morphological cues and disregarding the semantic class of the noun (Rodina, 2014; Rodina & Westergaard, 2012). Errors with stem-stressed neuter nouns also take longer to abate compared to more transparent cases (Janssen, 2016). However, the present research will leave such ambiguous cases beyond consideration.

4. Method

4.1 Participants

Fifty children participated in the experiment, including 21 Dutch-Russian bilinguals (age range: 5;0–6;7, mean age: 5;9; 14 females) and 29 Russian monolinguals (age range: 4;8–6;7, mean age: 5;8; 21 females). The bilingual children were recruited from a Russian complementary school in Amersfoort (central part of the Netherlands). These children were born and raised in the Netherlands, in most cases by a Russian-speaking mother and a Dutch-speaking father. They all attended a regular Dutch primary school (from age 4 onwards) and a Russian language school during the weekend. The children had no history of language impairment, as reported by teachers and parents. The monolingual participants were recruited from two pre-schools in Saint-Petersburg (Russia). All children were monolingual speakers of Russian and had age-appropriate language skills, normal motor, social-emotional and cognitive development, as reported by teachers. A group of adult L1-speakers of Russian (N=21, age range: 28–52, mean age: 41 years) was also tested to provide a reference point for child performance.

4.2 Materials

The experimental materials included 30 items. Each item consisted of two pictures of animals (for an example, see Figure 1), accompanied by two spoken sentences. In the first sentence the animals depicted in the pictures were mentioned as either the subject or the object (e.g. The fox met the duck). The subject of the second sentence was a pronoun (either *on* 'he' or *ona* 'she') referring to one of the animals in the first sentence (e.g. She wants a new toy).

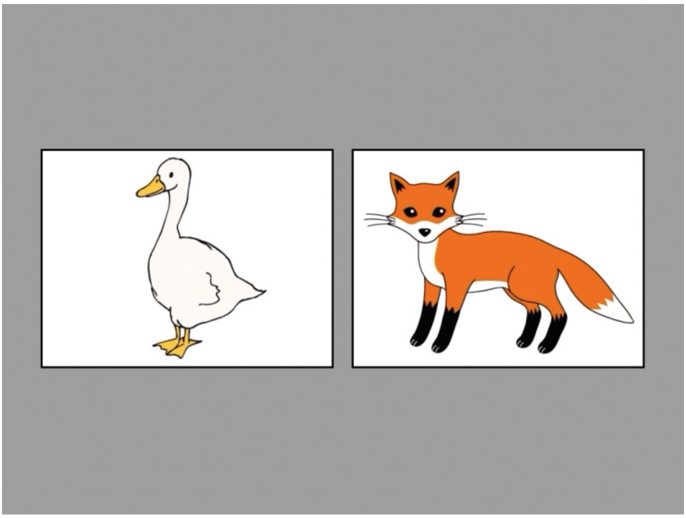

Figure 1. Example of a set of pictures used in the experiment

The 30 items were divided over three conditions (10 items per condition). In the ambiguous (same-gender) condition, the animal names had the same grammatical gender (they were either both masculine or both feminine). In these cases, the pronoun was not informative about the referent: grammatically, it could refer to either the subject or the object. In contrast, the other two conditions were informative (different-gender) conditions. In these cases, the pronoun referred to either the subject (subject condition) or the object (object condition) of the previous sentence. The example in Figure 1 is a different-gender trial because the noun *gus'* 'goose' is masculine and *lisa* 'fox' is feminine. Hence, the pronoun *ona* 'she' in this case unambiguously refers to the fox.

The sentences were recorded by a female native speaker of Russian. The sound files were manipulated in such a way that the onset of the pronoun was at 3.5 seconds after item onset, the onset of the next word (mostly the verb) was at 4.5 seconds after item onset. The total duration of the items was about 7 seconds. In half of the items the subject of the first sentence was presented on the left side of the screen, in the other half the subject of the sentence was presented on the right.

4.3 Apparatus

In the Netherlands, the experiment was run on a Tobii 1750 eye tracker, sampling at 50 Hz (every 20 ms). The items were presented on a 17-inch monitor via a computer running the Tobii's Clearview software. In Russia, the experiment was run on a Tobii T60 eye tracker, sampling at 60 Hz (every 16.6 ms) using Tobii Studio. The items were presented on a 17-inch monitor.

4.4 Procedure

The children were tested individually in a quiet room at their pre-school (monolinguals in St. Petersburg) or at the Russian complementary school (bilinguals in Amersfoort). The adults were tested in a quiet room in their homes or at the Russian complementary school. The participants did not have any task and were simply asked to look at the pictures on the screen and listen to the sentences. After the calibration, the experiment started. The whole experiment took about 5 minutes, including calibration.

4.5 Analysis

We analysed the data in the time window from the onset of the pronoun (at 3.5 seconds from item onset) until the end of the sentence, at 7 seconds from item onset. The data were analysed by means of a logistic regression, using the glmer function in R (Bates, Maechler, & Bolker, 2013). A logistic regression characterizes the data as binomial in that at a certain time point a subject can fixate on either picture, allowing us to assess the probability of looks to a picture over time. If the participant fixated elsewhere, the data were not included for analysis. Item and participant were included as random factors, Condition (ambiguous condition *vs.* subject condition *vs.* object condition), Time (in seconds) and Time^2 were included as fixed factors. The latter factor was included to test whether there was a curvilinear development in the proportion of looks at a picture. We also included the interaction of Condition and the two-time factors in the model. Random intercepts were included for participants and items.

If participants are sensitive to the gender of the pronoun, the percentage of looks at the target picture (either the subject or the object) should increase. In our analysis, the proportion of looks at the subject picture was the dependent variable. In the case of a pronoun referring to the subject, we expected this proportion to increase. In the case of a pronoun referring to the object, we expected this proportion to decrease. In the case of an ambiguous pronoun, other factors than gender marking determine the probability of looks at the target picture (e.g. subject bias, prosody). We expected the proportion of looks in this condition to be between those in the subject condition and the object condition.

5. Results

5.1 Adult participants

We first analysed the results in the adult group. Figure 2 presents the regression lines representing the probability of looks at the subject picture over the time course of the second clause. Since the first clause ends with the object, the probability of looking at the subject starts low. The lines in the picture show the development of the looks at the picture in the three conditions: When the pronoun refers to the subject there is an increase in looks at the subject picture, when the pronoun either refers to the object or is ambiguous, the probability of looks at the subject picture remains low. The model[1] for the adult data is presented in Table 2 (the ambiguous condition was taken as the baseline).

In the ambiguous condition, there was a linear effect of Time (parameter 2). The proportion of fixations on the subject picture increased after pronoun onset. In the subject condition the linear increase was much stronger (parameter 5). Also, there was a significant difference in the quadratic effect of Time: The increase was stronger at the beginning, and levelled off towards the end of the trial (parameter 6). In the object condition, there was a decrease in the proportion of fixations on the subject picture over Time compared to the ambiguous condition (parameter

1. The model specification in R was as follows:

(looksatsubject == 1)~group*condition + time*group*condition + I(time^2)*group* condition + (1|participant) + (1|item)

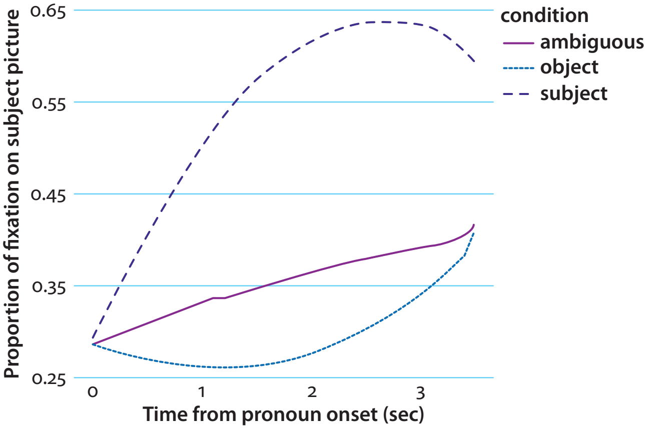

Figure 2. Probability of looks at the subject picture, by condition (adults)

8). Also, there was a quadratic effect of Time, indicating that the proportion of looks at the subject increased towards the end of the trial (parameter 9).

Table 2. Fixed and random effects from the model of the adult data

Fixed effects	*Estimate*	*SE*	*z-value*	*p*
1. Intercept	−0.93	0.18	−5.15	<.001
2. Time	2.70	1.08	2.50	.01
3. Time^2	−3.43	3.06	−1.12	.26
4. Subject	0.04	0.18	0.22	.82
5. Subject*Time	8.93	1.55	5.76	<.001
6. Subject* Time^2	−18.89	4.42	−4.28	<.001
7. Object	0.02	0.18	0.11	.91
8. Object*Time	−5.25	1.48	−3.55	<.001
9. Object* Time^2	14.71	4.18	3.52	<.001
Random effects	*Variance*	*SD*		
Participant	0.27	0.52		
Item	0.13	0.36		

5.2. Monolingual children

The probability of looks at the subject picture in the monolingual group is presented in Figure 3. The figure clearly shows that the children have a much weaker tendency to look at the picture that is referred to: The probability of looks at the picture denoting the subject of the first clause is around .50 across the whole second clause. The model for monolingual children is summarised in Table 3 (ambiguous condition is the baseline).

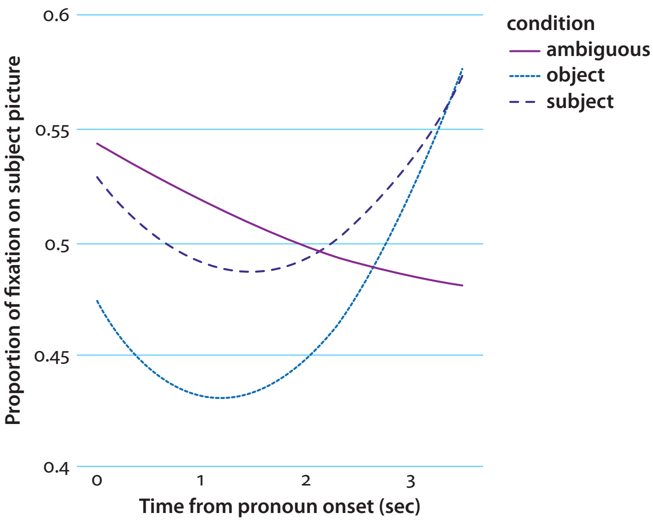

Figure 3. Probability of looks at the subject picture, by condition (monolingual children)

Table 3. Fixed and random effects from the model of the data of the monolingual children

Fixed effects	*Estimate*	*SE*	*z-value*	*p*
1. Intercept	0.18	0.13	1.40	.16
2. Time	−1.21	0.84	−1.45	.15
3. Time^2	1.32	2.31	0.57	.57
4. Subject	−0.04	0.16	−0.25	.80
5. Subject*Time	−1.26	1.18	−1.07	.28
6. Subject* Time^2	7.26	3.26	2.22	.03
7. Object	−0.32	0.16	−1.98	.046
8. Object*Time	−1.72	1.23	−1.40	.16
9. Object* Time^2	10.70	3.42	3.13	.002
Random effects	*Variance*	*SD*		
Participant	0.04	.21		
Item	0.12	0.35		

In the ambiguous condition, there was no effect of Time (parameters 2 and 3). In the subject condition, there was an increase in the proportion of looks at the subject, but only towards the end of the trial, as evidenced by the significant difference in the quadratic component (parameter 6). In the object condition, there was first a decrease in the proportion of looks at the subject, followed by an increase (parameter 9). In order to compare the object condition directly with the subject condition, we created an additional model with the subject condition as the baseline. In this model there was a difference between the subject and the object condition in the intercept only (Est = −0.28, *SE* = 0.07, *z* = −3.75, *p* < .001): The shape of the line was similar for the two conditions, but the overall proportion of looks at the subject picture was lower in the object condition.

5.3 Bilingual children

The probability of looks at the subject picture in the bilingual group is presented in Figure 4. As in the case of monolingual children, the probability of looking at the picture representing the subject of the first clause is close to .50 across the second clause. The model for bilingual children is summarised in Table 4 (ambiguous condition is the baseline).

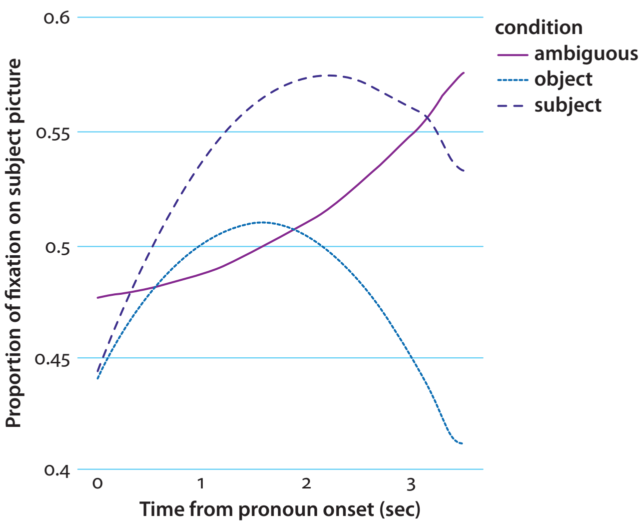

Figure 4. Probability of looks at the subject picture, by condition (bilingual children)

There were no effects in the ambiguous condition. In both the subject and the object condition there was an increase in the proportion of looks at the subject picture (parameters 5 and 8) followed by a decrease towards the end of the trial (parameters 6 and 9). An additional model with the subject condition as the baseline showed that there was no difference between the subject and object condition across the analysis window (0–3.5 seconds after pronoun onset).

Visual inspection of the data shows that there might be a difference between the subject and the object condition towards the end of the trial. To analyse this,

Table 4. Fixed and random effects from the model of the data of the bilingual children

Fixed Effects	**Estimate**	***SE***	***z*-value**	***p***
1. Intercept	−0.10	0.12	−0.85	.40
2. Time	0.05	1.09	0.05	.96
3. Time^2	3.04	3.17	0.95	.34
4. Subject	−0.11	0.14	−0.78	.44
5. Subject*Time	4.79	1.51	3.17	.002
6. Subject* Time^2	−13.92	4.36	−3.19	.001
7. Object	−0.12	0.15	−0.85	.39
8. Object*Time	3.58	1.64	2.19	.03
9. Object* Time^2	−14.61	4.75	−3.08	.002
Random Effects:	*Variance*	*SD*		
Participant	0.06	0.24		
Item	0.07	0.26		

we computed a new model over the second part of the analysis window (1.5–3.5 seconds after pronoun onset); the subject condition was taken as a baseline. In this model the object condition differed from the subject condition in the intercept (Est = 3.06, $SE = 0.76$, $z = 4.04$, $p < .001$): There were more looks at the subject picture in the subject condition. There was also a significant effect of Time (Est = −26.48, $SE = 6.16$, $z = -4.30$, $p < .001$): There was a decrease in the proportion of fixations on the subject picture, but this was only significant in the object condition. Finally, there was a difference in the quadratic effect of Time (Est = 48.99, $SE = 12.24$, $z = 4.00$, $p < .001$): After an initial rise in the proportion of looks at the subject picture, there was a decrease in both conditions, but the decrease was stronger in the object condition.

5.4 Summary of the results

The adult participants were clearly sensitive to gender cues in pronoun processing. Upon hearing the informative pronoun, the proportion of looks to the subject picture increased in the subject condition and decreased in the object condition. Both informative (different-gender) conditions differed from the uninformative (ambiguous condition), but in different directions. Monolingual children were also sensitive to pronoun gender. In this group, the proportion of looks at the subject picture was significantly higher in the subject condition than in the object condition.

The results in the bilingual group were less straightforward. Unlike the monolingual groups, the bilingual participants did not show sensitivity to pronoun gender in the time window from 0 to 3.5 seconds after item onset. However, in the second part of this time window (1.5–3.5 seconds after pronoun onset), the pattern of their looks in the subject condition differed from the object condition: In that time window the bilingual children looked more at the picture of the subject when pronoun gender was compatible with the gender of the subject antecedent. In this time window, the proportion of looks at the subject picture also rose faster and decreased less steeply than in the object condition. All in all, these results suggest that Dutch-Russian bilinguals are sensitive to pronoun gender and use it in online processing of coherent discourse. However, the effect is observed later than in monolingual children and adults, so bilingual children are slower than monolinguals in processing gender information.

6. Discussion

Prior research has shown that the acquisition of grammatical gender is demanding for children with reduced exposure to the target language, as evidenced by frequent gender agreement errors in production. Tribushinina and Mak (2016) report that 5-year-old Dutch-Russian bilinguals dominant in Dutch still perform at chance in the production of Russian pronominal gender, whereas Russian monolinguals make very few errors already at age 3. The present paper aimed to shed more light on the underlying causes of these errors. We hypothesized that if bilinguals' poor performance in production is due to incomplete acquisition of grammatical gender (lacking grammatical representations), they should not be sensitive to gender cues in online processing either. However, if the frequent production errors are merely a reflection of performance limitations (or in Nicoladis' terms "epiphenomena of speech production"), bilingual children should use gender cues when processing a coherent discourse.

The results of this study are more compatible with the latter scenario. Bilingual children, like their monolingual peers and Russian-speaking adults, were able to use gender information in the online processing of anaphoric pronouns. In the informative condition, in which the anaphoric pronoun could be unambiguously related either to the subject or to the object of the previous clause (based on the grammatical gender of the antecedent), the proportion of looks at the subject picture was higher when the pronoun referred to the subject of the previous clause and lower when the pronoun referred to the object. However, the bilingual group was slower than the monolingual comparison groups in orienting towards the target picture: The bilingual group only showed evidence of gender processing after 1.5 seconds from pronoun onset. The finding that bilinguals were less efficient (slower) in online processing of pronouns is consistent with earlier research demonstrating that bilinguals have a reduced processing capacity because they have to deal with two competing language systems (Bialystok, 2007; Paradis, 2010; Roberts, Gullberg, & Indefrey, 2008; Sorace & Serratrice, 2009; White, 2011).

The results of this study are also in line with earlier research demonstrating comprehension–production asymmetries in bilinguals across different linguistic domains (Ågren & Van de Weijer, 2013; Chondrogianni & Marinis, 2012; Chondrogianni, Marinis et al. 2015; Chondrogianni, Vasić, 2015; Gagarina & Sauermann, in press; Mak et al., 2017) and, more specifically, in the domain of gender agreement in the attributive domain (Blom & Vasić, 2011; Brouwer et al., 2008; Unsworth, 2013). The current results extend these findings to gender agreement in the pronominal domain. The question arises as to why bilingual children make so many errors in the production of grammatical gender, if they do have relevant abstract representations. Given the performance in the receptive tasks, problems in production are likely to be a result of the processing cost of dealing with two language systems.

One processing disadvantage of bilingualism is that bilinguals usually have slower lexical access than monolinguals, even if the size of their receptive vocabularies is comparable to that of monolingual peers (Yan & Nicoladis, 2009). Selection of lemmas from the mental lexicon leads to partial activation of words from both languages, including their grammatical features. Speech production is a very fast, automatic process, and slower and/or less efficient lexical access may lead to breakdowns in accessing the relevant morpho-syntactic features resulting in speech production errors (Chondrogianni, Marinis et al., 2015; Chondrogianni, Vasić et al., 2015; VanPatten, Keating, & Leeser, 2012; White, 2003, 2011).

Another explanation of target-like comprehension despite variable production is also related to the processing cost in bilingualism, but in this case specifically to the competition between the typologically different language systems in a bilingual mind (Nicoladis, 2006; Sorace & Serratrice, 2009). There is plenty of evidence of systematic cross-linguistic influence in the speech of bilingual individuals (Argyri & Sorace, 2007; Döpke, 1998, 2000; Hulk & Müller, 2000; Müller & Hulk, 2001; Serratrice, 2007; Serratrice, Sorace, & Paoli, 2004). It is especially hard to inhibit the dominant language (operationalized here as a language which children use most), because a dominant language has a lower activation threshold due to its more frequent use (Argyri & Sorace, 2007; Tribushinina et al., 2017). There is evidence from production studies that bilingual children sometimes make gender errors that are compatible with the gender of the counterpart noun in their other language (Cantone & Müller, 2008). For one, all pronoun gender errors attested in the speech of Dutch-Russian children in Tribushinina and Mak (2016) were compatible with the typological properties of Dutch: Bilinguals over-used the masculine pronoun *on* 'he' where the feminine pronoun *ona* 'she' should have been used (even though feminine pronouns are more phonologically salient and similar to masculine pronouns in terms of frequency). In Dutch *hij* 'he' is a default pronoun used with reference to all individual objects and animals, irrespective of the grammatical gender of the noun. Hence, it is plausible to assume that cross-linguistic transfer from Dutch (i.e. the children's dominant language) is the prime cause of production errors.

It is also possible that the slower response to gender cues attested in the current study is not only due to general computational pressures in bilingual processing, but also due to cross-linguistic influence from Dutch. Recall that in the bilingual group, unlike in the two monolingual groups, there was an initial increase in the proportion of looks to the subject picture, irrespective of the pronoun gender. This initial response is similar to the subject bias in pronoun interpretation in Dutch (Kaiser, 2011; Roberts et al., 2008). It is plausible to assume that bilinguals initially resort to a default strategy from their dominant language (to relate ambiguous pronouns to subjects), because they have less solid representations of noun gender or because they have difficulties accessing gender information in their mental lexicon (see Janssen, Meir, Baker, & Armon-Lotem, 2015 for similar findings regarding the processing of case and word-order cues by Russian-Dutch and Russian-Hebrew bilinguals). This would explain why the gender effect in the bilingual group was observed later than in monolingual children and adults. Even though the pattern of the current results is compatible with this explanation, it should be treated with caution because we have not tested pronoun resolution in the children's dominant language (Dutch). Furthermore, we assumed that all our participants were dominant in Dutch because they were raised in the Netherlands by a Russian mother and a Dutch father, and went to a regular Dutch school, but we did not measure the exact amount of their exposure to each language. Future research should test bilingual participants in both languages and relate their performance in each language to the amount of exposure in that language.

Another promising avenue for future research would be to test bilingual children whose two languages have partly overlapping and partly different systems of grammatical gender. For example, in both German and Russian the word for 'ostrich' is masculine (congruent gender), whereas the word for 'dog' is masculine in German and feminine in Russian (incongruent gender). If bilingual German-Russian children perform worse on incongruent trials than on congruent trials, this would be more rigorous evidence of cross-linguistic influence in gender processing. Our group is currently exploring this possibility.

Acknowledgements

We would like to thank an anonymous reviewer for the helpful comments. We thank all children, parents and teachers who have made this investigation possible. We are particularly grateful to Elvira Skiba, Viktorija Golovkina, Elizaveta Andreiushina, Katja Becker, Tatjana Erfurt and Ekaterina Abrosova for their help with recruitment and testing.

Funding

This research was supported by a Marie Curie International Research Staff Exchange Scheme Fellowship within the 7th European Community Framework Programme (grant number 269173).

References

Argyri, E., & Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism: Language and Cognition*, 10, 79–99. https://doi.org/10.1017/S1366728906002835

Arnold, J., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, 22(4), 527–565. https://doi.org/10.1080/01690960600845950

Ågren, M., & Van de Weijer, J. (2013). Number problems in monolingual and bilingual French-speaking children. *Language, Interaction and Acquisition*, 4(1), 25–50. https://doi.org/10.1075/lia.4.1.02agr

Bates, D., Maechler, M., & Bolker, B. (2013). *Lme4: Linear mixed-effects models using S4 classes* (R package version 0.999999-2). Retrieved from <http://CRAN.R-project.org/package=lme4> (3 May, 2020).

Bialystok, E. (2007). Cognitive effects of bilingualism: How linguistic experience leads to cognitive change. *International Journal of Bilingual Education and Bilingualism*, 10, 210–223. https://doi.org/10.2167/beb441.0

Blom, E., Polišenská, D., & Weerman, F. (2008). Articles, adjectives and age of onset: the acquisition of Dutch grammatical gender. *Second Language Research*, 24, 289–323.

Blom, E., & Vasic, N. (2011). The production and processing of determiner-noun agreement in child L2 Dutch. *Linguistic Approaches to Bilingualism*, 1(3), 265–290. https://doi.org/10.1075/lab.1.3.03blo

Brouwer, S., Cornips, L., & Hulk, A. (2008). Misrepresentation of Dutch neuter gender in older bilingual children? In E. Gavrusheva & B. Haznedar (Eds.), *Trends in child second language acquisition* (pp. 83–96). Amsterdam: John Benjamins. https://doi.org/10.1075/lald.46.06bro

Brouwer, S., Sprenger, S., & Unsworth, S. (2017). Processing grammatical gender in Dutch: Evidence from eye movements. *Journal of Experimental Child Psychology*, 159, 50–65. https://doi.org/10.1016/j.jecp.2017.01.007

Cantone, K. F., & Müller N. (2008). Un nase or una nase? What gender marking within switched DPs reveals about the architecture of the bilingual language faculty. *Lingua*, 118, 810–826. https://doi.org/10.1016/j.lingua.2007.05.007

Chondrogianni, V., & Marinis, T. (2012). Production and processing asymmetries in the acquisition of tense morphology by sequential bilingual children. *Bilingualism: Language and Cognition*, 15, 5–21. https://doi.org/10.1017/S1366728911000368

Chondrogianni, V., Marinis, T., Edwards, S., & Blom, E. (2015). Production and on-line comprehension of definite articles and clitic pronouns by Greek sequential bilingual children and monolingual children with specific language impairment. *Applied Psycholinguistics*, 36, 1155–1191. https://doi.org/10.1017/S0142716414000101

Chondrogianni, V., Vasić, N., Marinis, T., & Blom, E. (2015). Production and on-line comprehension of definiteness in English and Dutch by monolingual and sequential bilingual children. *Second Language Research*, 31(3), 309–341. https://doi.org/10.1177/0267658314564461

Corbett, G. G., & Fraser, N. M. (2000). Default gender. In B. Unterbeck (Ed.), *Gender in grammar and cognition* (pp. 55–91). Berlin: Mouton. https://doi.org/10.1515/9783110802603.55

Cornips, L., & Hulk, A. C. J. (2006). External and internal factors in bilingual and bidialectal language development: Grammatical gender of the Dutch definite determiner. In C. Lefebvre, L. White, & C. Jourdan (Eds.), *L2 acquisition and creole genesis* (pp. 355–378). Amsterdam: John Benjamins. https://doi.org/10.1075/lald.42.21cor

Döpke, S. (1998). Competing language structures: The acquisition of verb placement by bilingual German-English children. *Journal of Child Language*, 25, 555–584. https://doi.org/10.1017/S0305000998003584

Döpke, S. (2000). Generation of and retraction from cross-linguistically motivated structures in bilingual first language acquisition. *Bilingualism: Language and Cognition*, 3, 209–226. https://doi.org/10.1017/S1366728900000341

Fhlannchadha, S. N., & Hickey, T. M. (2017). Acquiring an opaque gender system in Irish, an endangered indigenous language. *First Language*, 37(5). https://doi.org/10.1177/0142723717702942

Foursha-Stevenson, C., & Nicoladis, E. (2011). Early emergence of syntactic awareness and cross-linguistic influence in bilingual children's judgments. *International Journal of Bilingualism*, 15, 521–534. https://doi.org/10.1177/1367006911425818

Franceschina, F. (2005). *Fossilized second-language grammars: The acquisition of grammatical gender*. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.38

Gagarina, N., & Sauermann, A. (In press). Patterns of anaphora resolution in Russian: Bilingual and monolingual settings. *Linguistic Approaches to Bilingualism*.

Gvozdev, A. N. (1961). *Formirovanie u rebeka grammatičeskogo stroja russkogo jazyka* [*Language development of a Russian child*]. Moscow: APN RSFSR.

Honselaar, W. (2014). De frequentie van flexieklassen van het substantivum in het Russisch [Frequency of inflectional classes of Russian nouns]. In W. Honselaar, J. Stelleman, & W. G. Weststijn (Eds.), *To the point. Festschrift for Eric de Haard* (pp. 119–142). Amsterdam: Pegagus.

Huettig, F., & Brouwer, S. (2015). Delayed anticipatory spoken language processing in adults with dyslexia: Evidence from eye-tracking. *Dyslexia*, 21, 97–122. https://doi.org/10.1002/dys.1497

Hulk, A. C. J., & Cornips, L. (2006). Neuter gender and interface vulnerability in child L2/2L1 Dutch. In S. Unsworth, T. Parodi, A. Sorace, & M. Young-Scholten (Eds.), *Paths of development in L1 and L2 acquisition* (pp. 107–134). Amsterdam: John Benjamins. https://doi.org/10.1075/lald.39.06hul

Hulk, A., & Müller, N. (2000). Crosslinguistic influence in bilingual children: Object omission and root infinitives. In C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *BUCLD 24: Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 546–557). Sommerville, MA: Cascadilla Press.

Janssen, B. (2016). *The acquisition of gender and case in Polish and Russian*. Amsterdam: Pegasus.

Janssen, B., Meir, N., Baker, A., & Armon-Lotem, S. (2015). On-line comprehension of Russian case cues in monolingual Russian and bilingual Russian-Dutch and Russian-Hebrew children. In E. Grillo & K. Jepson (Eds.), *BUCLD 39: Proceedings of the 39th Annual Boston University Conference on Language Development* (pp. 266–278). Somerville, MA: Cascadilla Press.

Kaiser, E. (2011). Salience and contrast effects in reference resolution: The interpretation of Dutch pronouns and demonstratives. *Language and Cognitive Processes*, 26(10), 1587–1624. https://doi.org/10.1080/01690965.2010.522915

Keij, B., Cornips, L., Van Hout, R., Hulk, A., & Van Emmerik, J. (2012). The acquisition of grammatical gender and the definite determiner in Dutch by L1-TD, L1-SLI, and eL2 children. *Linguistic Approaches to Bilingualism*, 2(4), 379–403. https://doi.org/10.1075/lab.2.4.02kei

Kupisch, T., Müller, N., & Cantone, K. F. (2002). Gender in monolingual and bilingual first language acquisition: Comparing Italian and French. *Lingue e Linguaggio*, 1, 107–147.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18(3), 193–198. https://doi.org/10.1111/j.1467-9280.2007.01871.x

Mak, W. M., Tribushinina, E., Lomako, J., Gagarina, N., Abrosova, E., & Sanders. T. (2017). Connective processing by bilingual children and monolinguals with specific language impairment: Distinct profiles. *Journal of Child Language*, 44, 329–345. https://doi.org/10.1017/S0305000915000860

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 139–162). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.27.09mar

Müller, N., & Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition*, 4, 1–21. https://doi.org/10.1017/S1366728901000116

Nicoladis, E. (2006). Cross-linguistic transfer in adjective-noun strings by preschool bilingual children. *Bilingualism: Language and Cognition*, 9(1), 15–32. https://doi.org/10.1017/S136672890500235X

Orgassa, A., & Weerman, F. (2008). Dutch gender in specific language impairment and second language acquisition. *Second Language Research*, 24(3), 333–364. https://doi.org/10.1177/0267658308090184

Paradis, J. (2010). The interface between bilingual development and specific language impairment. *Applied Psycholinguistics*, 31, 227–252. https://doi.org/10.1017/S0142716409990373

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition*, 30, 333–357. https://doi.org/10.1017/S0272263108080480

Rodina, Y. (2008). *Semantics and morphology: The acquisition of grammatical gender in Russian* (Unpublished doctoral dissertation). University of Tromsø.

Rodina, Y. (2014). Variation in the input: Child and caregiver in the acquisition of grammatical gender in Russian. *Language Sciences*, 43, 116–132. https://doi.org/10.1016/j.langsci.2013.10.007

Rodina, Y., & Westergaard, M. (2012). A cue-based approach to the acquisition of grammatical gender in Russian. *Journal of Child Language*, 39(5), 1077–1106. https://doi.org/10.1017/S0305000911000419

Rodina, Y., & Westergaard, M. (2017). Grammatical gender in bilingual Norwegian–Russian acquisition: The role of input and transparency. *Bilingualism: Language and Cognition*, 20, 197–214. https://doi.org/10.1017/S1366728915000668

Sanford, A. J., & Filik, R. (2007). "They" as a gender-unspecified singular pronoun: Eye-tracking reveals a processing cost. *The Quarterly Journal of Experimental Psychology*, 60(2), 171–178. https://doi.org/10.1080/17470210600973390

Serratrice, L. (2007). Cross-linguistic influence in the interpretation of anaphoric and cataphoric pronouns in English-Italian bilingual children. *Bilingualism: Language and Cognition*, 10, 225–238. https://doi.org/10.1017/S1366728907003045

Serratrice, L., Sorace, A., & Paoli, S. (2004). Crosslinguistic influence at the syntaxpragmatic interface: Subjects and objects in English-Italian bilingual and monolingual acquisition. *Bilingualism: Language and Cognition*, 7, 183–205. https://doi.org/10.1017/S1366728904001610

Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, 13, 195–210. https://doi.org/10.1177/1367006909339810

Thomas, E. M., & Gathercole, V. (2007). Children's productive command of grammatical gender and mutation in Welsh: An alternative to rule-based learning. *First Language*, 27, 251–278. https://doi.org/10.1177/0142723707077056

Tribushinina, E., & Mak, W. M. (2016). Ce que les corpus de production orale ne peuvent montrer - apports de l'oculométrie (eye-tracking) dans la recherche sur le bilinguisme et sur la dysphasie. In M. Kilani-Schoch, C. Surcouf, & A. Xanthos (Eds.), *Nouvelles technologies et standards méthodologiques en linguistique* (pp. 161–196). Lausanne: Université de Lausanne.

Tribushinina, E., Mak, W. M., Andreiushina, E., Dubinkina, E., & Sanders, T. (2017). Connective use by bilinguals and monolinguals with SLI. *Bilingualism: Language and Cognition*, 20(1), 98–113. https://doi.org/10.1017/S1366728915000577

Unsworth, S. (2008). Age and input in the acquisition of grammatical gender in Dutch. *Second Language Research*, 24, 365–396. https://doi.org/10.1177/0267658308090185

Unsworth, S. (2013). Assessing the role of current and CUMULATIVE exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, 16(1), 86–110. https://doi.org/10.1017/S1366728912000284

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, 35, 765–805. https://doi.org/10.1017/S0142716412000574

VanPatten, B., Keating, G. D, & Leeser, M. J. (2012). Missing verbal inflections as a representational problem. *Linguistic Approaches to Bilingualism*, 2(2), 109–140. https://doi.org/10.1075/lab.2.2.01pat

White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511815065

White, L. (2011). Second language acquisition at the interfaces. *Lingua*, 121(4), 577–590. https://doi.org/10.1016/j.lingua.2010.05.005

Yan, S., & Nicoladis, E. (2009). Finding *le mot juste*: Differences between bilingual and monolingual children's lexical access in comprehension and production. *Bilingualism: Language and Cognition*, 12(3), 323–335. https://doi.org/10.1017/S1366728909990101

Zufferey, S., Mak, W., Degand, L., & Sanders, T. (2015). Advanced learners' comprehension of discourse connectives: The role of L1 transfer across on-line and off-line tasks. *Second Language Research*, 31(3), 389–411. https://doi.org/10.1177/0267658315573349

Language impairment

A syntactically based treatment of relative clauses

Three case studies of Italian children with cochlear implant

Silvia D'Ortenzio[1], Silvia Montino[2], Alessandro Martini[2], Patrizia Trevisi[2] and Francesca Volpato[1]
[1]Ca' Foscari University of Venice / [2]University of Padua

In this paper, we describe three case studies of a syntactic intervention given to three Italian-speaking children with cochlear implants (CIs) to improve proficiency in the production and comprehension of relative clauses, and also their narrative skills. The methodology adopted for the syntactic intervention follows previous studies on aphasic patients with agrammatism and children with developmental language disorders (DLD). Indeed, these studies have shown that an explicit teaching of syntactic rules helps in the recovery and improvement of complex structures derived by syntactic movement. Results showed a general improvement in the production and comprehension of relative clauses, and generalization effects to untrained structures. Moreover, also narrative skills improved. Results were maintained several months after the end of the syntactic intervention.

Keywords: language acquisition, relative clauses, narrative skills, syntactic intervention, hearing loss, cochlear implants

1. Introduction

According to the World Health Organisation (WHO), hearing loss is one of the most common disabilities of human beings. Indeed, it affects more than 466 million people worldwide (6,1% of the World's population) (WHO, 2018).

Hearing loss consists in an inability to hear sounds. It can affect one or more parts of the ear; therefore, it is possible to distinguish three types of hearing loss: conductive hearing loss (CHL); sensorineural hearing loss (SHL), and mixed

https://doi.org/10.1075/lald.64.09dor

hearing loss (MHL). CHL is caused by damage in the sound-conducting apparatus.[1] SHL results from damage or malfunctioning of the cochlea or of the auditory nerve. MHL origins from damage in the sound-conducting apparatus plus damage of the cochlea or auditory nerve (Martini et al., 2013).

Since hearing loss causes a decrease in the quantity and the quality of the linguistic input, deaf children show problems in the acquisition of an oral language. According to Friedmann and Szterman (2006, 2011), as reported by Penke and Wimmer (2018), one of the causes of difficulties in language acquisition for children with hearing loss is a syntactic deficit relating to the syntactic movement of the object to the left periphery of the sentence, thus causing misprocessing of syntactically complex sentences.

When the hearing loss is higher than 70 dB, a cochlear implant (CI) may be prescribed (Vincenti et al., 2014). The CI is an electronic artificial sensory organ which directly stimulates the residual fibres of the acoustic nerve, which transfers auditory perceptions to the cortical areas in the central nervous system (Guida et al., 2014). The CI is composed of two parts: the external part and the internal part. The external part is called the *speech processor* and is composed of a microphone that catches the sounds, a processor that turns sounds into signals, and a transmitter. The transmitter communicates with the internal part called the *cochlear implant*. The internal part converts the signals into electric energy and sends them to the electrode array positioned in the cochlea, thus stimulating the nerve fibres in the cochlea; the brain recognises the signals as sounds (Martini et al., 2013).

Although children with CIs may show good proficiency with some linguistic aspects, such as vocabulary acquisition and/or speech perception (Caselli et al., 2012; Rinaldi & Caselli, 2013, among many others), some of them still show difficulties with complex syntactic structures, such as relative clauses (Friedmann & Szterman, 2006; Volpato, 2010, 2012, among many others), *wh*-questions (Szterman & Friedmann, 2015; Penke & Wimmer, 2018; D'Ortenzio & Volpato, 2020, among many others) and clitic pronouns (Guasti et al., 2014).

Relative clauses are problematic not only for children with hearing loss and fitted with hearing aids (HAs) and/or CIs (De Villiers, 1988; Friedmann et al., 2008; Volpato & Vernice, 2014, among many others), but also for several populations with language disorders, such as patients with agrammatic aphasia (Grillo, 2008; Garraffa & Grillo, 2008), children with developmental language disorders (DLD) (Friedmann & Novogrodsky, 2007; Contemori & Garraffa, 2010), and children with developmental dyslexia (Pivi et al., 2016; Delage & Durrleman, 2018).

1. The sound-conducting apparatus involves the ear canal, the ear drum, the middle ear and its little bones (malleus, incus, and stapes).

Several studies have pointed out that it is possible to improve the individual's competence in the comprehension and production of syntactically complex sentences through the explicit teaching of syntactic rules, including very young children (Roth, 1984), patients with agrammatic aphasia (Thompson, 2003; Thompson & Shapiro, 2005), children with DLD (Ebbels et al., 2007; Levy & Friedmann, 2009), bilingual children (Volpato & Bozzolan, 2017; Piccoli, 2018), L2 speakers (De Nichilo, 2017), and adolescents with developmental dyslexia (Piccoli, 2018). This methodology has been found effective because it improves both the participants' production and the comprehension of syntactically complex sentences, such as relative clauses. Moreover, the explicit teaching of syntactic rules forces the participant to think about his/her language consciously, allowing him/her to retell what s/he has learned (Ellis, 2009).

Considering the good results observed with several populations across different languages, the aim of this paper is to describe three case studies on the syntactic intervention given to three Italian children with CIs.

The paper is organised as follows. First, the acquisition of relative clauses in Italian and how these structures are processed by Italian-speaking children with CIs will be presented. Then, previous studies on the treatment of relative clauses will be briefly introduced. After that, the methodology adopted during these three studies and the results collected before treatment will be described. We then move to the description of the protocol of the syntactic intervention given to children with CIs. Then, the data collected at the end of the syntactic intervenion will be discussed. Finally, a short discussion and conclusion will be provided at the end of the paper.

2. The acquisition of relative clauses in Italian

Relative clauses are subordinate clauses which modify a nominal element known as the antecedent.

In restrictive relative clauses the antecedent is modified by limiting the number of referents for it. Italian restrictive relative clauses are introduced by the complementizer *che* 'that'.

Relative clauses are derived by movement of the subject or the object NP towards a non-argument position in Spec-CP (A′ movement, Kayne, 1994; Bianchi, 1999), leaving a gap in the position in which the NP is interpreted. The participants of our study were tested on subject relative clauses (SRs) (1), object relative clauses with an embedded preverbal subject (ORs) (2), and object relative clauses with an embedded postverbal subject (ORps) (3).

(1) La giraffa che <la giraffa> pettina i gatti. SR
the giraffe that <the giraffe> combs the cats
'The giraffe that combs the cats.'

(2) I gatti che la giraffa pettina <i gatti>. OR
the cats that the giraffe combs <the cats>
'The cats that the giraffe combs.'

(3) I gatti che pettina la giraffa <i gatti>. ORp
the cats that combs the giraffe <the cats>
'The cats that the giraffe combs.'

Relative clauses involve the formation of a long-distance dependency between the relative 'head' and the gap in the original position. This relationship is short in SRs and long in ORs. Across different populations, subject relatives are easier to produce and comprehend than object relatives, and object relatives with a preverbal subject are easier than object relatives with a postverbal subject. For Italian, this result has been pointed out by several studies in different populations, such as typically developing children and adults (Guasti & Cardinaletti, 2003; Belletti & Contemori, 2010; Volpato, 2010), children with DLD (Contemori & Garraffa, 2010), children with CIs (Volpato, 2012; Volpato & Vernice, 2014), and patients with agrammatic aphasia (Garraffa & Grillo, 2008).

Italian-speaking typically developing children between 3;0 and 3;11 produce 61% of correct SRs. This percentage raises to 90% at the age of 4;0 (Belletti & Contemori, 2010) and is higher than 90% at adolescence and adulthood (Carpenedo, 2011; Volpato, 2010). The production of ORs is more problematic than SRs; typically developing children between the age of 5;3 and 7;5 produce 18% of ORs, while adolescents and adults do not produce ORs (Volpato, 2010). Indeed, typically developing children resort to several strategies in order to avoid the production of an OR. For example, they produce ORs with resumptive clitic pronouns as in (4), or they produce different types of passive object relatives, such as copular (5) or causative (6) (Volpato, 2010; Manetti & Belletti, 2013, among many others).

(4) I gatti che la giraffa li pettina.
the cats that the giraffe them.CL comb.3SG
'The cats that the giraffe combs them.'

(5) I gatti che vengono pettinati dalla giraffa.
the cats that come.3PL comb.PAST.PART.PL by-the giraffe
'The cats that are combed by the giraffe.'

(6) I gatti che si fanno pettinare dalla giraffa.
the cats that themselves make.3PL comb.INF by-the giraffe
'The cats that had themselves combed by the giraffe.'

As for comprehension, typically developing children comprehend SRs already between the age of 3;4 and 3;11, when they reach the 91% of correctness (Adani, 2011). Conversely, at the same age, typically developing children show a low comprehension of ORs (53%) and of ORps (36%). This percentage increases at the age of 7;0 when typically developing children comprehend 89% of ORs and 70% of ORps. An adult-like level is reached by the age of 11, when the performance is above chance level for all types of structure: SRs, ORs, ORp (Arosio et al., 2009).

In this paper, both production and comprehension of relative clauses are investigated. According to some researchers, these two tasks present an asymmetry, namely, children acquire first the competence in language production and then they acquire the competence in language comprehension (Hendricks & Koster, 2010). This asymmetry has been found in the acquisition of reflexive pronouns (Matthews et al., 2009; Spenader et al., 2009, among many others), and canonical word order (Subject-Verb-Object) (Chapman & Miller, 1975; McClellan et al., 1986). Several studies have reported the same trend for the acquisition of relative clauses (Håkansson & Hansson, 2000; Guasti, 2017, among many others). Children start to produce relative clauses at around the age of 3; while they only fully comprehend relative clauses at around the age of 6. However, regardless of the age at which children start producing and comprehending relative clauses, the sentence production can be more demanding than comprehension due to the involvement of several processes in the former task such as the retrieval of words and meaning from memory (Gennari et al., 2009). Therefore, it is possible that the percentage of correct responses in production tests will be lower the percentages of correct responses in comprehension tests.

Like typically developing children, children with CIs also show the subject/object asymmetry in the acquisition of relative clauses. Volpato and Vernice (2014) analysed the production of relative clauses in a group of 13 Italian-speaking children with CIs, aged 7;9–10;8 (mean age: 9;2). Results showed that children with CIs produced 88% of SRs and 23% of ORs. ORs were mostly replaced by ambiguous sentences (17%) (*Mi piacciono i gatti che pettinano le giraffe* 'I like the cats that combs the giraffe') and passive relatives (26%) (*Mi piacciono i gatti che sono pettinati dalla giraffa* 'I like the cats that are combed by the giraffe'). Children with CIs resorted also to other strategies, such as incomplete or ungrammatical sentences, or sentences with Theta-role inversion, which emerge to a smaller extent in the production of normal hearing peers. A further outcome of this study is that older children, from both the experimental and control groups, resorted to passive relatives when an OR was elicited, showing age-appropriate language acquisition. Volpato and Adani (2009) analysed the comprehension of 8 Italian-speaking children with CIs aged 6;9–9;3 (mean age: 7;9). Children with CIs showed the same trend as the age- and language-matched normal hearing children, namely

SRs were performed better than ORs and ORps, and ORs were performed better than ORps. However, children with CIs obtained lower scores than children with normal hearing (SR: 89%; ORs: 55%, and ORps: 22%).

3. Treatment of relative clauses

As pointed out by many studies on agrammatic patients (Thompson, 2003; Thompson & Shapiro, 2005;), children with DLD (Ebbels, 2007; Levy & Friedmann, 2009, among many others), children with CI (D'Ortenzio et al., 2017; D'Ortenzio 2018), Italian L2 speakers (Volpato & Bozzolan, 2017; De Nichilo, 2017; Piccoli, 2018), and adolescents with developmental dyslexia (Piccoli, 2018), complex structures derived by *wh*-movement, like relative clauses, can be rehabilitated through explicit teaching of syntactic rules or principles, such as verb argument structure (Chomsky, 1981), the Theta Criterion (Chomsky, 1981), and *wh*-movement (Chomsky, 1977). Thompson (2003) has shown that explicit teaching of syntactic rules also provides generalization effects to untrained structures derived by the same syntactic movement. In more detail, generalization occurs to untrained structures that are less complex than the trained structures. For example, the production of *wh*-questions can improve after a syntactic intervention focused on cleft sentences, but cleft sentences cannot improve after a training focused on *wh*-questions. Moreover, Ballard and Thomson (1999) and Jacobs and Thompson (2000) showed that generalization does not extend to structures derived by a syntactic rule other than the trained one. For instance, the use of passive sentences does not improve after a linguistic training focused on *wh*-questions. Indeed, the former structure is derived by A-movement, while the latter is derived by A′-movement. Thompson and Shapiro (2005) have reported that linguistic training of syntactically complex structures can also lead to improvements in discourse patterns and narrative skills (increase in mean length of utterances (MLU), increase in the number of grammatical sentences, and increase in the number of verbs with the correct selection of arguments).

In a recent study carried out with a child with CIs, it was found that the treatment of relative clauses based on the explicit teaching of syntactic rules is a valid method to improve the production and comprehension of relative clauses also in this population (D'Ortenzio et al., 2017, D'Ortenzio, 2018).

4. Methodology

This section provides information related to the tests used in all the three studies, a description of the participants' personal and clinical data, and a description of the syntactic intervention.[2]

Participants

Three case studies are presented in this paper.

The participants are three Italian-speaking children with hearing loss and fitted with CIs. They were all born in hearing families and they were trained orally; they do not use or know a sign language. The participants were all selected and tested in the Ear Nose Throat Clinic, Department of Neurosciences, University of Padua.

LB is an 8;5-year-old male, who was first diagnosed with hearing impairment at the age of 2 and promptly received HAs. He received a CI at the age of 2;7. At the time of the syntactic intervention, LB had experience of CI use for 5;10 years. LB received speech and language therapy once a week. He was followed by two assistant teachers for five hours a day at school, and by a communication assistant for twelve hours a week. His performance was compared to the performance of a small control group composed of two normal hearing Italian-speaking children aged 8;3 and 8;8 (TD1). The experiment with this participant is presented in D'Ortenzio et al. (2017).

ES is a 10;5-year-old female, who was diagnosed with hearing impairment at the age of 1;10. After she was fitted with HAs, she was only exposed to oral language. ES received the CI at the age of 8;4. At the time of the experiment, she gained auditory experience of 1;11 years with her CI. ES wears the CI on the right ear, and the HA on the left ear. At the time of treatment, ES was receiving speech-treatment once a week, she was followed by an assistant teacher at school and by an educator eight hours a week. Her performance was compared to the performance of a small control group composed of three normal hearing Italian-speaking children aged between 10;3 and 10;11 (TD2).

MM is a 9;9-year-old female, suffering from bilateral severe-to-profound sensorineural hearing loss and fitted with bilateral CIs. She was diagnosed and received her first HA at the age of 5 months. Since she did not receive enough benefit from her HAs, she received a CI at the age of 2;9 on the right ear. When she was 9;8 years old, she received the second CI on the left ear. She attended speech and language therapy once or twice a week. MM's performance was compared to the

2. The tasks assessing production and comprehension of relative clauses were administered to all three children with CI, while the Frog Story was administered only to ES and MM.

performance of a small control group composed of three normal hearing Italian-speaking children aged between 9;5 and 9;7 (TD3).

Materials

The elicited production task

The elicited production task developed by Volpato (2010) follows the approach by Friedmann and Szterman (2006) and is used to analyse the production of SRs and ORs. The task aimed at forcing the participant to produce subject and object relative clauses. The task consists of 24 experimental trials (12 questions eliciting a SR, 12 questions eliciting an OR). The experimenter shows two pictures to the participant, and asks the participant to express a preference between the two options. The two items used for the elicitation of a SR are presented in Figure 1; those aimed at eliciting an OR are presented in Figure 2.

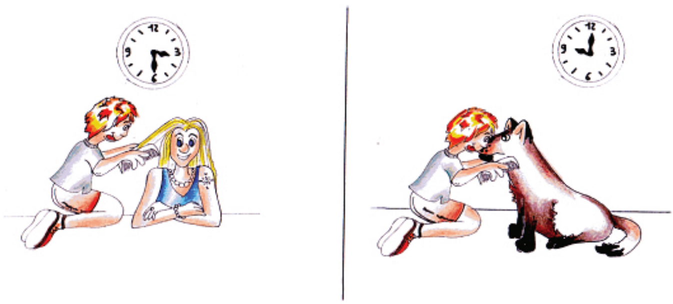

Figure 1. Item eliciting the production of a SR

Elicitation of a SR:

Experimenter:	*Ci sono due immagini. Nella prima il bambino pettina la mamma, nella seconda il bambino pettina il cane. Quale bambino ti piace di più? Inizia con: "Mi piace il bambino…"* 'There are two pictures. In the first, the child combs the mother; in the second, the child combs the dog. Which child do you like? Start with "I like the child…"'
Target:	*Mi piace il bambino che pettina la mamma/il cane.* 'I like the child that combs the mother/the dog.'

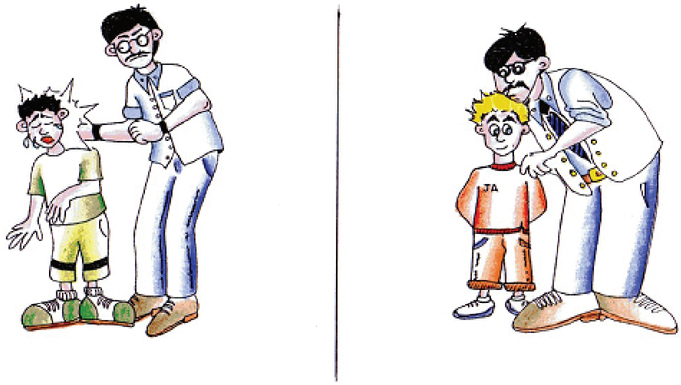

Figure 2. Item eliciting the production of an OR

Elicitation of an OR:

Experimenter: *Ci sono due immagini. Nella prima il papà picchia il bambino, nella seconda il papà bacia il bambino. Quale bambino ti piace di più? Inizia con: "Mi piace il bambino…".*
'There are two pictures. In the first the father hits the child; in the second the father kisses the child. Which child do you like? Start with: "I like the child…"'

Target: *Mi piace il bambino che il papà picchia/bacia.*
'I like the child that the father hits/kisses'.

The experimental sentences were semantically reversible, namely they contained verbs whose theta roles could be compatible with both DPs in the clause, preventing the child from deriving the meaning of the sentence by relying on semantic or pragmatic cues. All verbs were transitive and in the present tense to avoid difficulties related to the presence of auxiliaries and past participle morphology, which are often problematic for children with hearing loss and may increase the difficulty of the task. (Chesi, 2006)

The comprehension task

The comprehension task was developed by Volpato (2010). This task aimed at investigating the comprehension of SRs, ORs, and ORps. After the participant has heard a sentence read by the experimenter, he/she had to touch the character matching the sentence. For each trial, two different scenes were presented to the participant: in the first scene, some characters perform an action (e.g., the giraffe greets the bunny), and in the second scene, the action is the same, but the theta roles are reversed (e.g., the bunny greets the giraffe). The task consisted of 48 experimental trials and 12 ambiguous sentences, i.e. sentences that could be interpreted as either a SR (*Tocca la giraffa che saluta il coniglio*, 'Touch the giraffe that greets the bunny') or an ORp (*Tocca il coniglio che saluta la giraffa*, 'Touch the rabbit that the giraffe greets'). Only reversible transitive verbs were used. Again, verbs were in the present tense. Figure 3 shows a pictures included in the comprehension task.

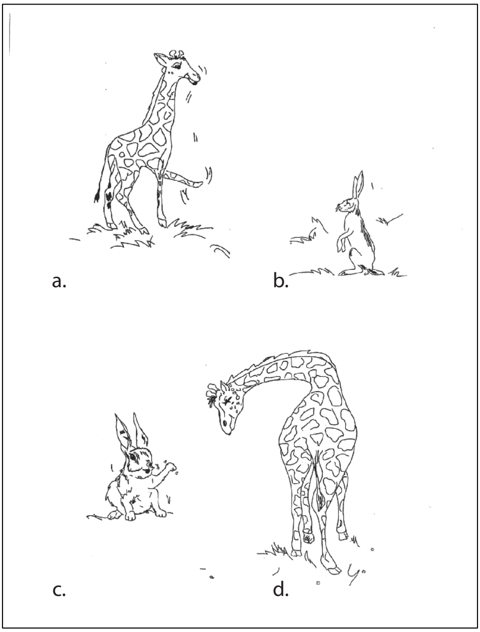

Figure 3. Example of a picture included in the comprehension task

The Frog story

The well-known Frog story ("Frog, where are you?", Meyer, 1969) is a book composed of 24 pictures telling the story of a boy, a dog, and a frog. The task consists in telling the story described by the pictures. In a nutshell, the boy owns a dog and a frog, but the frog escapes during the night. When the boy and the dog wake up the morning after, they notice that the frog has escaped, so they start to search for it first in the house, and then outside in a wood, where they find several animals. At the end of the story, the boy and the dog find the frog. The Frog story helps therapists to analyse whether the child has an appropriate use of complex sentences, has an adequate Mean Length of Utterance (MLU), makes morphosyntactic errors, and displays narrative competence. This task was developed for children ranging in age from 3;6 to 12. Storytelling requires the activation of several discourse-specific skills, that involve distinct types of competence. Therefore, narrative production represents a crucial tool through which it is possible to study and analyse children's typical and atypical development of cognitive, linguistic, pragmatic, and social understanding in different observational settings (D'Amico et al., 2008). Starting from this assumption, two of the participants (ES and MM) were assessed by means of the Frog story to analyse their narrative skills. LB was not assessed using the Frog story. Indeed, the aim of the first case-study was to investigate the fruitfulness of the exercises in the training of the target structures, without further exploring generalization effects to other structures and/or narrative skills.

5. Results prior to intervention

LB's performance

LB's experiment was presented in D'Ortenzio et al. (2017) and D'Ortenzio (2018). LB was tested with the elicited production task and the comprehension task before and after the syntactic intervention, and also five months after the end of the syntactic intervention.

Before the syntactic intevention, LB showed the typical asymmetry between subject and object relatives in both the production and the comprehension tasks. The results of the elicited production task showed a preserved production of SRs (83%) and a problematic production of ORs (0%), i.e. he avoided the target structure and produced ungrammatical sentences, which had the structure: e.g. *Mi piace il bambino che fa il diritto al cane di seguirlo* 'I like the child that makes the right to the dog to follow him'. The TD1 group showed better performance than LB in the production of SRs. They also avoided the production of targeted ORs, but, differently from LB, they produced passive relatives, i.e. *Mi piace il bambino che è inseguito dal cane* 'I like the child that is followed by the dog', an age-appropriate structure (see Section 1 above). Figure 4 shows the percentage of correct answers in the production task.

However, LB showed good comprehension of relative clauses. The percentage of correct responses was similar to those of the control children. In SRs and ORs,

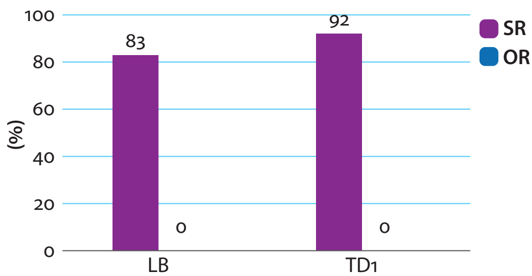

Figure 4. Percent correct by LB (before syntactic intervention) and the TD1 group in the elicited production task

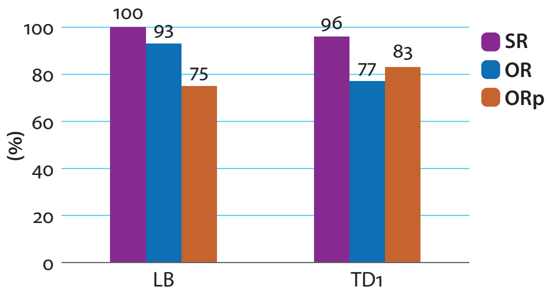

Figure 5. Percent correct in the comprehension task by LB (before syntactic intervention) and the TD1 group

percentages are slightly higher for LB than for the control participants; the opposite was found for ORps. Figure 5 shows the percentage of correct responses in the comprehension task by LB and his control group (TD1).

Comparing comprehension and production, performance is very similar in both tasks, although both LB and his controls showed slightly higher percentages of accuracy in the comprehension task than in the production task.

ES's performance

As mentioned above, the assessment of narrative skills was introduced during this second experiment. Therefore, ES was tested with the elicited production task, the comprehension task, and the Frog story before and after the syntactic intervention, and also two months after the end of the intervention in order to analyse whether the effects of the syntactic intervention had been maintained.

Figure 6 shows ES's responses in the elicited production task compared with her control group.

ES's performance before the syntactic intervention showed the typical asymmetry between subject and object relatives. SRs were correctly produced while

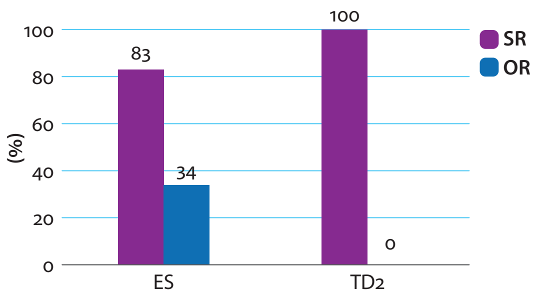

Figure 6. Percent correct by ES (before syntactic intervention) and the T2 group in the production task

ORs were problematic. When an OR was elicited, ES sometimes produced target ORs, with either gaps (*Mi piacciono i bambini che il papà sta pettinando* 'I like the children that the father is combing') or resumptive full DPs (*Mi piace il bambino che il papà sta lavando il bambino* 'I like the child that the father is washing the child'). She also resorted to other strategies to avoid the production of ORs, such as Theta role inversion (*Mi piace il bambino che abbraccia la mamma* 'I like the child that hugs the mother', instead of *Mi piace il bambino che la mamma abbraccia* 'I like the child that the mother hugs') or the use of verbs different from the target ones. Her normal hearing age peers produced 100% SRs and no ORs, which were substituted with passive sentences.

ES showed most of her difficulties in the comprehension task. Figure 7 shows ES's results compared to those of the control group.

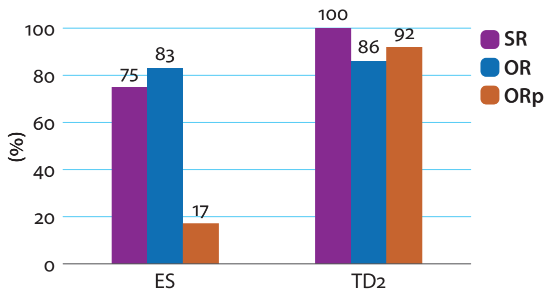

Figure 7. Percent correct in the comprehension task by ES (before syntactic intervention) and the TD1 group

Indeed, she showed lower percententage in the comprehension of SRs (75%) and ORps (17%) than ORs (83%). As expected, the control participants comprehended SRs at ceiling, while the comprehension of ORs and ORps was found to be slightly lower.

ES showed worse performance than her control group in both production and comprehension of relative clauses.

MM's performance

Differently from LB and ES, MM's linguistic competence was evaluated twice before the syntactic intervention in order to provide a baseline. Indeed, several studies suggest a baseline period consisting in assessing participants on the same task/structure at least twice before the intervention starts. This helps to assess possible progress without the intervention (Thompson & Shapiro, 2005, 2007; Ebbels, 2014, 2017). Changes during the baseline could be caused by language acquisition or by practice effects of the test (Ebbels, 2014).

During the first assessment of production before treatment (PRE1), MM showed the typical asymmetry between SRs (67%) and ORs (17%). Moreover, she produced an ungrammatical structure in order to avoid the production of SRs (*Mi piacciono i bambini alla quale guardano i cavalli* 'I like the children to which watch the horses'), and ORs (*Mi piacciono i bambini alla quale vengono baciati dai nonni* 'I like the children to which are kissed by the grandparents'). However, when she was retested the second time before we started the syntactic intervention (PRE2), she never resorted to this strategy; rather she produced passive relatives instead of ORs, thus showing the same pattern as her typically developing age peers (*Mi piacciono i bambini che vengono baciati dai nonni* 'I like the children that are kissed by the grandparents'). Finally, the production of SRs also increased during the PRE2 assessment. Figure 8 shows MM's performance in the production task compared to her control group.

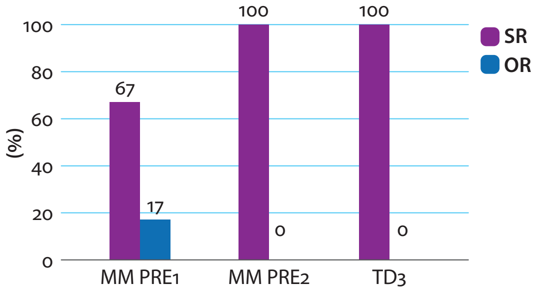

Figure 8. Percent correct in the elicited production task related to MM's baseline period (PRE1 and PRE2) and her control group (TD3)

As for comprehension, during the PRE1 and PRE2 assessments, MM showed very high accuracy on all three sentence types, in contrast to the control participants, who showed the typical asymmetry between SRs, ORs, and ORps, as shown in Figure 9.

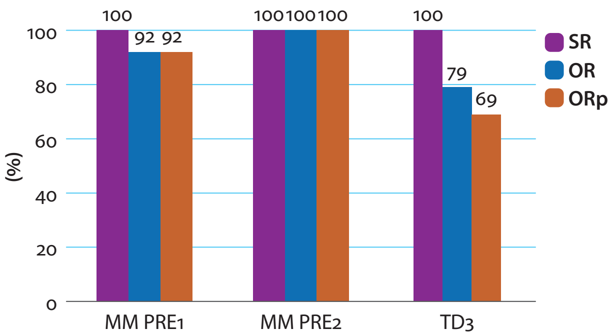

Figure 9. Percent correct in the comprehension task related to MM's baseline period (PRE1 and PRE2) and her control group (TD3)

Before the linguistic training MM showed very good performance, even better than her hearing age peers. Nevertheless, she was selected for the linguistic training to analyse possible generalization effects to untrained structures and narrative skills.

6. Explicit intervention of relative clauses in Italian-speaking children with CIs

The treatment of relative clauses was based on the explicit teaching of some syntactic rules. As mentioned above, the syntactic intervention addressed to LB has been already described in D'Ortenzio et al. (2017) and D'Ortenzio (2018). In this first attempt at training relative clauses, SRs, ORs, and ORps were the target structures for the training, without considering the possibility of generalization effects to untrained structures. In the next studies, those involving ES and MM, we decided to train only ORs in order to control generalization effects on both easier structures (SRs), and more difficult structures (ORps), and on narrative skills (Frog story). Even though ORs are only occasionally produced by children and adults, they were chosen for the syntactic intervention since they are derived by a very complex movement, namely the object moves from the position where it is generated to a new position crossing the subject. This choice follows the so-called "complexity effect", namely the treatment is grounded on more complex structures in order to help generalization effects to simpler untrained structures derived by the same syntactic movement (Thompson, 2003).

The syntactic interventions we carried out lasted between two and three months and comprised between six and seven lessons, in three different phases. The first focused on verb argument structure and the Theta criterion; the second focused on *wh*-movement; the third and last phase was dedicated to the review of the material taught during the syntactic intervention.

Only when the participants showed that they had reached good mastery of the material considered during a session or a phase, did they progress to the next session/phase (Ebbels, 2007; Levy & Friedmann, 2009).

In the following sections we provide the protocol used during the syntactic interventions addressed to ES and MM.

Phase 1: Verb argument structure and Theta criterion

This stage comprised two or three sessions during which the experimenter explained verb argument structure and the Theta criterion to the participant so as to turn his/her implicit knowledge into explicit knowledge, to be used for support during the explanation of *wh*-movement (Levy & Friedmann, 2009). Every session started with some exercises followed by a teaching part. Giving the exercises before the explanation of the syntactic rules helps the participant to start thinking about his/her language so as to internalize some issues, rather than just providing conventional rules.

This session started with a simple exercise consisting in giving a list of several verbs with which the participant should write a simple sentence. After this first activity, the participant was asked to underline the subject of the sentence with a blue colour, and the object of the sentence with an orange colour. Then, the experimenter put emphasis on the similarities and differences between verbs, namely, verbs always have a subject, but not always an object; therefore, there are different types of verbs depending on the number of elements that accompany them. Thereafter, the experimenter explained the Theta criterion to the participants. Following Chomsky (1981), every argument in the sentence must receive one and only one thematic role, and the verb must assign a thematic role to each argument. Therefore, a verb like 'follow' must assign agent and theme roles to its arguments, and these roles must be assigned in each sentence in which the verb 'follow' appears. Taking into consideration the sentence structure in Italian, it was explained to participants that verbs generally assign the agent role to the NP that occupies the suject position before it, and the theme role to the NP that follows it in object position. In order not to confuse the participants, only the agent and the theme roles were considered during the explanation of the Theta criterion. After the teaching part, the experimenter asked the participants to find the agent and the theme roles in several sentences and to underline them with different colours.

To make the verb argument structure and the Theta criterion clearer, the experimenter introduced a metaphor in which the verb is compared to an orchestra leader. Indeed, the verb is like an orchestra leader who chooses how many musicians must play music. Orchestra leaders are different, they can choose only one musician (monovalent verbs), two musicians (bivalent verbs), or three musicians (trivalent verbs). After the orchestra leader has chosen one or more musicians, he also decides which instruments the musicians must play and assigns a different musical instrument to each musician.

Phase 2: Wh-movement

This phase consisted of three or four sessions. This phase of the syntactic intervention was focused on *wh*-movement in ORs.[3] Each session started with a review of the topics taught during the previous sessions to strengthen the participant's

3. In our first attempt, the participant was taught syntactic movement in SRs, ORs, and ORps following the approach adopted by Levy and Friedmann (2009).

knowledge of verb argument structure and the Theta criterion. The review was carried out by working with several exercises, such as a grammaticality judgment task, in which the participant was asked to decide whether a sentence was grammatical or not. Whenever the answer was negative, he/she had to explain the reason why the sentence was ungrammatical.

Following previous studies (Ebbels et al., 2007; Levy & Friedmann, 2009), *wh*-movement was explained by resorting to a card game, composed of several cards, representing the different elements of the sentence. As shown in Figure 10, white cards represent NPs, yellow cards represent verbs, green cards represent clauses external to the relative clause, red cards represent the complementizer, and blue cards represent the trace. As in previous studies (D'Ortenzio et al., 2017; Volpato & Bozzolan, 2017; D'Ortenzio, 2018), the first step in explaining *wh*-movement consisted in writing a simple sentence with the constituents in canonical order.[4] Then, the experimenter added an external clause, which was necessary to introduce the relative clause. Once the object is moved to the left, the experimenter leaves a trace-card in the position from which the object has moved. Finally, the complementizer is inserted in the new sentence and the experimenter ties the trace-card with the noun-card acting as the object of the sentence with a ribbon. Figure 10 shows the several steps to derive an ORs.

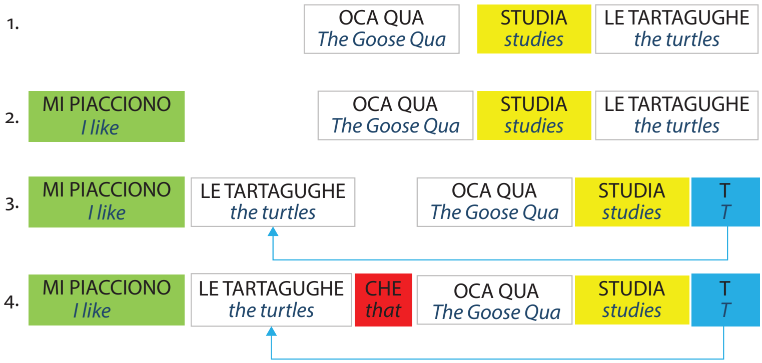

Figure 10. the four steps useful for the derivation of an OR

In order to avoid the card game becoming an automated exercise, with the participant moving the object of the sentence without reflecting on the movement, we developed a "syntactic Scrabble". This game allows the participants to combine cards in order to create simple sentences and relative clauses. Like a conventional board game, the "syntactic Scrabble" has a grid on which the cards are positioned. Participants have several cards to start the game and then they get points

4. The canonical order of constituents in Italian is Subject-Verb-Object (SVO).

depending on the difficulty of the sentence, for example the creation of simple sentences gives 1 point, while relative clauses, which are more complex, carry a weight of 3 points. The higher the number of points collected by the participant, the more new cards can be bought to create new sentences. The participant with the highest score wins the game.

Phase 3: Review

The last stage of the syntactic intervention comprised only one session and was focused on the general review of the topics taught during the intervention. As in the previous sessions, also in this last session the participants were asked to complete exercises on verb argument structure and the Theta Criterion, i.e. they were given a story and they were asked to find verbs and to assign them to one category (monovalent, bivalent, trivalent). Then, they were asked to find the agent and theme roles for each verb and to underline the theta roles with different colours.

7. Results after intervention

This section presents the results of the data collected at the end of the linguistic training. LB and ES were tested before and after the syntactic intervention, and also several months later. As described above, MM was tested twice before the beginning of the syntactic intervention. She was not re-tested several months after the end of the treatment.[5] Thus, it was not possible to investigate whether the effects of the linguistic training were maintained in the longer term.

Firstly, we will present for each participant the results in the elicited production task and in the comprehension task, and then for ES and MM also the results of the assessment of narrative skills.[6]

LB's performance

Immediately after the end of the syntactic intervention the results showed that LB performed at ceiling in each task and in each structure and did not show any asymmetry in the production and comprehension of relative clauses. Figure 11

5. MM was not retested because she got too close to the experimenter and she did not stay focused on the tasks during the assessment.

6. The *Frog story* was administered to ES and MM. MM's narrative skills were assessed also with a different story so as to control generalization effects better.

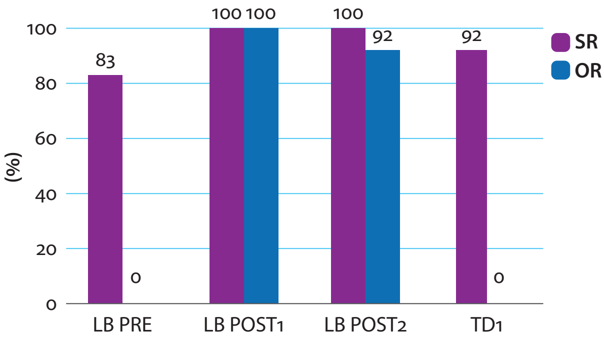

Figure 11. Percent correct in the elicited production task comparing LB's performance before (PRE) and after (POST1, POST2) the syntactic intervention with the performance of a small control group

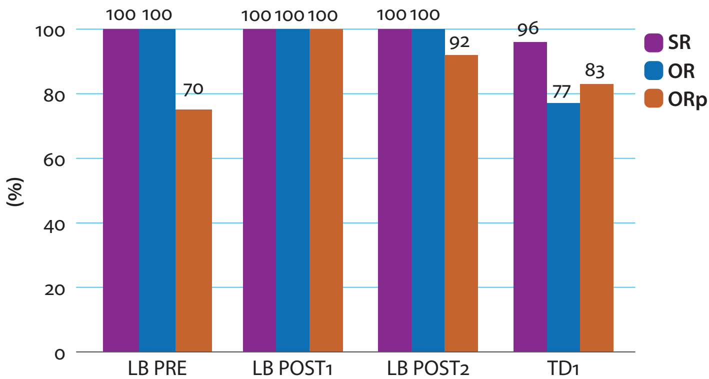

Figure 12. Percent correct in the comprehension task comparing LB's performance before (PRE) and after (POST1, POST2) the syntactic intervention with the performance of a small control group

and Figure 12 show the percentage of correct responses in the elicited production task and in the comprehension task, respectively.

LB's performance was compared with that of normal hearing age peers. The controls show high scores in the production of SRs, while ORs were replaced with passive sentences, which are an age-appropriate strategy in order to avoid the production of ORs. The control participants showed the typical asymmetry between SRs and ORs in the comprehension test, and the percentage of correct answers provided are somewhat lower than those of LB. The results after treatment suggest that LB has reached a syntactic awareness that allows him to analyse and interpret relative clauses correctly.

LB was also tested five months after the end of the syntactic intervention, and his performance showed no regression; he still performed at ceiling. It is worth mentioning that in the POST2 assessment, LB resorted to a different strategy to produce ORs, namely, he produced ORs with resumptive clitic pronouns, like *Mi piace il bambino che l'orso lo morde* 'I like the child that the bear bites him'. This piece of data will be discussed in the discussion section below.

ES's performance

After the syntactic intervention, ES's performance showed improvement in trained (ORs) and untrained structures (SRs and ORps) and in narrative skills, thus showing generalization effects. Moreover, the effects of the syntactic intervention were maintained two months after the completion of the treatment. Before intervention, ES showed the typical asymmetry between SRs and ORs in the production test. This asymmetry was found in the comprehension of relative clauses on the pretest but only with ORps. The data collected soon after the syntactic intervention showed improved performance on both tasks compared to performance before treatment. Results were maintained and improved over time. As discussed in Section 4, ES's controls performed in line with their age, namely, they produced a high number of correct SRs and they replaced the production of ORs with the passive sentences. Moreover, their performance during the comprehension task showed the typical asymmetry between SRs and ORs, namely, the former structure was comprehended better than the latter. Figure 13 shows the percentage of correct responses of ES compared to her control group in the elicited production task, while Figure 14 shows the results of the comprehension task.

Looking at the results, we may assume that ES's improvement is the result of the linguistic training and a gained syntactic awareness of the syntactic rules that are involved in the derivation of relative clauses. Furthermore, ES was able

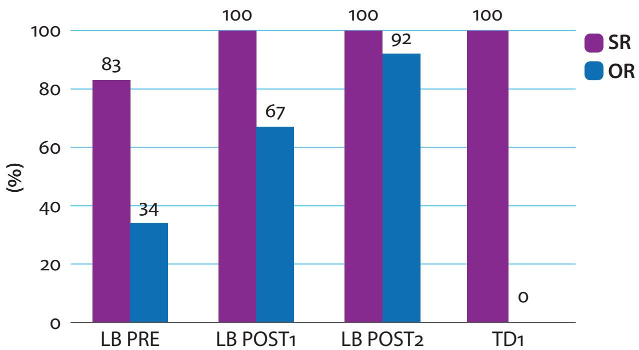

Figure 13. Percent correct in the elicited production task comparing ES's performance before (PRE) and after (POST1, POST2) the syntactic intervention with the performance of a small control group

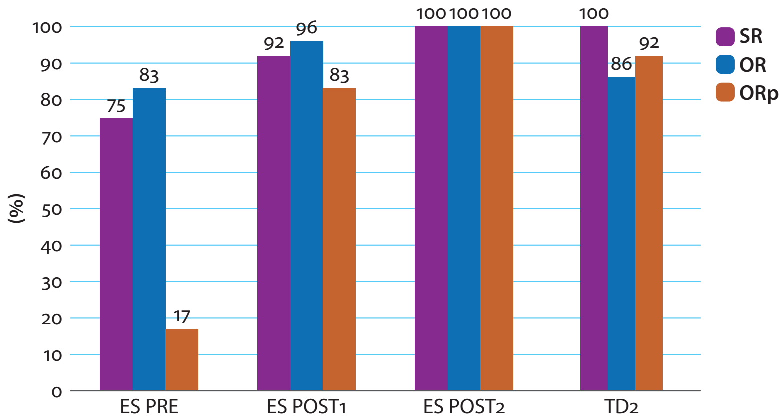

Figure 14. Percent correct in the comprehension task comparing ES's performance before (PRE) and after (POST1, POST2) the syntactic intervention with the performance of a small control group

to generalize beyond the structure that the treatment focused on, namely OR, to both SR and Orp.

MM's performance

Figure 15 and Figure 16 show MM's performance in the production and in the comprehension of relative clauses, respectively.

MM's results after treatment do not show any improvement between pretest 2 and the post test in either production or comprehension of relative clauses, since the percentage of accuracy was very high before the beginning of the syntactic intervention. Interestingly, like her normal hearing age peers, MM always produced passive relatives instead of ORs. Unlike the controls, MM showed no asymmetry in the comprehension of SRs and ORs/ORps.

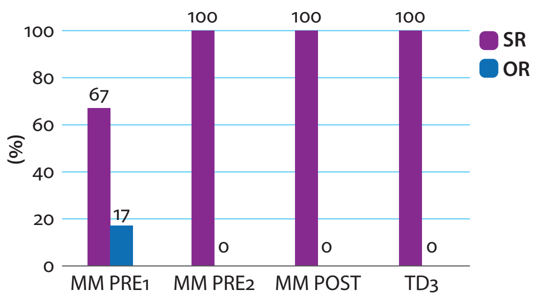

Figure 15. Percent correct in the elicited prodution task comparing MM's performance before (PRE1, PRE2) and after (POST) the syntactic intervention with the performance of a small control group

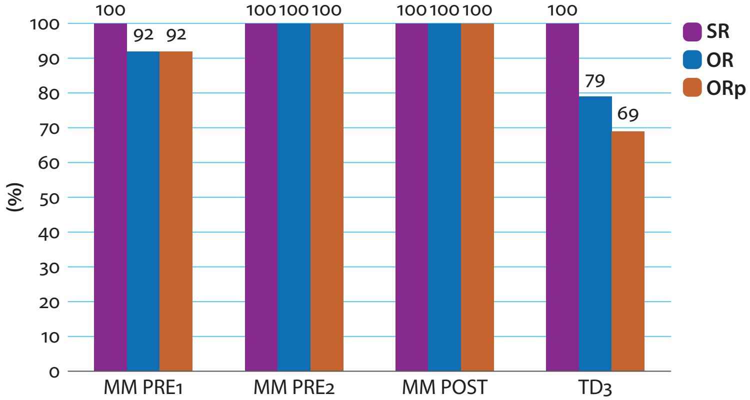

Figure 16. Percent correct in the comprehension task comparing MM's performance before (PRE1, PRE2) and after (POST) the syntactic intervention with the performance of a small control group

ES and MM's improvement of narrative skills

As mentioned above, only ES and MM were also assessed on their narrative skills. Indeed, this type of assessment was introduced with these two participants in order to investigate the possible generalization effects to narrative skills, as pointed out by Thompson and Shapiro (2005). The Frog story was administered to both ES and MM before and after the syntactic intervention. Moreover, at the end of the intervention, MM's narrative skills were assessed also with a new story for a better control of generalization effects.

Table 1 shows participant's performances in the assessment of narrative skills.

Focusing on ES's performance, after the intervention based on the explicit teaching of syntactic rules she showed a decrease in the number of hesitations during the storytelling, and a decrease in the production of ungrammatical or incomplete sentences.

As for MM, she showed good narrative skills already before the syntactic intervention. However, after the intervention, she showed a low number of hesitations in both the Frog Story and the new story; in addition, in the new story, she produced a lower number of ungrammatical or incomplete sentences than in the Frog Story. Moreover, MM showed an increased number of subordinate clauses.

The decreased number of ungrammatical and incomplete sentences and the increase of the production of subordinate clauses may be attributable to the explicit teaching of syntactic rules, as also suggested by Thompson and Shapiro (2005) for patients with agrammatic aphasia. The fact that the number of hesitations was reduced after the linguistic intervention is considered a sign of improvement by

Table 1. Analysis of the narrative skills of ES and MM before and after treatment.

		ES				MM					
		FROG STORY				FROG STORY				NEW STORY	
		PRE		POST		PRE		POST		POST	
		N	Rate	N	Rate	N	Rate	N	Rate	N	Rate
	Words	274		261		415		349		421	
	Sentences	36		53		74		66		82	
	MLU	4.89		4.92		5.5		5.3		5	
	Hesitations	19		6		17		4		5	
Sentences	Main	10/56	18%	15/53	28%	29/74	39%	24/66	36%	38/82	46%
	Coordinate	23/56	41%	21/53	40%	21/74	28%	16/66	24%	25/82	30%
	Subordinate	10/56	18%	9/53	17%	7/74	9%	14/66	21%	11/82	13%
	Relative	1/56	2%	0/53	–	5/74	7%	4/66	6%	5/82	6%
	Passive	0/56	–	0/53	–	0/74	–	1/66	2%	0/82	–
	Ungrammatical/Uncomplete	12/56	21%	8/53	15%	12/74	16%	7/66	11%	3/82	4%

many speech and language therapists since the participant shows a more homogeneous speech pattern, which is due to continuous work on complex sentences.

8. Discussion

In this study, we have presented the results of syntactic interventions based on explicit teaching of syntactic rules, addressed to three Italian-speaking children with CIs (LB; ES; MM). All children were tested before and after linguistic intervention. On the one hand, children were selected for this study because they showed difficulties in the production and comprehension of relative clauses (albeit to different extents), and especially in those structures with non-canonical word order, i.e., ORs and ORps. On the other hand, they were chosen to investigate whether the syntactic intervention focused on a specific syntactic structure (the relative clause) may lead to further improvement also in narrative skills.

The syntactic intervention described in this paper followed the therapeutic protocols administered to people with agrammatic aphasia by Thompson and Shapiro (1995, 2005; see also Thompson, 2003), and the syntactic intervention provided to a 12-year-old Hebrew-speaking child with syntactic DLD by Levy and Friedmann (2009). Syntactic intervention was based on the explicit teaching of verb argument structure, the Theta Criterion, and *wh*-movement through the use of metaphors (the verb compared to an orchestra leader), written and oral exercises (identifying the theta roles in different types of sentences), and games (the 'syntactic scrabble'). Despite the common approach adopted for intervention, the three studies are different from one another. While LB's intervention focused on the training of SRs, ORs, and ORps (D'Ortenzio et al., 2017; D'Ortenzio, 2018), ES and MM were trained only on ORs in order to investigate generalization effects to untrained structures (SR and ORps) and narrative skills. In addition, for MM, we also collected baseline data with two measurements before treatment. Baseline should be considered as a goal in case study designs, and some studies have pointed out the importance of making at least two measurements until the baseline is stable (Thompson & Shapiro, 2005, 2007; Ebbels, 2014, 2017), in order to investigate the efficacy of treatment. In addition to this, the protocol presented here differentiates from the studies carried out with patients with agrammatic aphasia and children with DLD. Indeed, while the experiments carried out by Thompson and Shapiro (1995, 2005), and Levy and Friedmann (2009), lasted over six months and consisted of a high number of sessions, the intervention protocol presented here consists of only seven sessions and lasted three months.

The first case study has already been discussed by D'Ortenzio et al. (2017) and D'Ortenzio (2018), who proved for the first time for Italian the effectiveness of syntactic intervention in a child with CI. Indeed, after a six-sessions linguistic training based on the explicit teaching of *wh*-movement in SRs, ORs, and ORps, LB showed ceiling performance in both comprehension and production of all trained structures. Moreover, in the comprehension task (Volpato, 2010), he was able to identify ambiguous sentences, which could be interpreted as either SRs or ORps (*Tocca i pinguini che spingono le giraffe* 'Touch the penguins that push the giraffes' / 'Touch the penguins that the giraffes push'). LB's performance was also assessed five months after the completion of the syntactic intervention. LB's production of SRs still showed ceiling effects, while the production of ORs included some sentences containing a gap in the object position (*Mi piace il bambino che il papà pettina* 'I like the child that the father combs') and some containing a resumptive clitic pronoun (*Mi piace il bambino che il papà lo pettina* 'I like the child that the father combs him'). Resumptive clitic pronouns in relative clauses are frequent in the production of younger normal hearing children (Guasti & Cardinaletti, 2003; Volpato, 2010), and in colloquial language (Cinque, 1988). Crucially, LB no longer produced ungrammatical sentences.

The second and third case studies, addressed to ES and MM, respectively, aimed at analysing generalization effects to untrained structures and narrative skills. For both participants, only ORs were trained.

Before the syntactic intervention, ES correctly produced both SRs and target ORs, although the performance slightly differed from that of the control children.

As for comprehension, percentages of SRs and ORps were lower than those of controls, probably because of a delay in the acquisition of these syntactically complex structures, presumably caused by later intervention with CI. The low comprehension of ORps is likely to be due to the unexpected word order of constituents in this type of sentence. Such an explanation was also suggested by other studies investigating relative clauses in children with CI (Volpato & Adani, 2009; Volpato & Vernice, 2014). ES's narrative skills were inadequate for her age and, even though she produced some complex structures, her oral production was characterized by ungrammatical sentences, incorrect theta-role assignment, and low MLU. After a syntactic intervention comprising seven sessions and lasting three months, both trained (ORs) and untrained structures (SRs, ORps) improved. As in previous studies (Thompson & Shapiro, 1995, 2005), improvement was also observed in her narrative skills, with a decrease in the number of ungrammatical sentences and correct use of agreement. Two months after the completion of the linguistic training, ES's performance showed further improvement in both production and comprehension of relative clauses.

The third case study is concerned with the linguistic intervention given to MM. Even though MM showed good competence in almost all the structures analysed during the pre-training assessment, she was selected for this study because her parents were interested in the experiment. The participant was assessed twice before the start of the syntactic intervention, in order to assess the rate of possible progress without the intervention, and to measure the treatment efficacy. To assess the baseline performance only relative clauses were administered. At the first assessment before intervention, MM produced ungrammatical structures in order to avoid production of SRs and ORs. However, she never resorted to this strategy during the second assessment before the syntactic intervention, during which she produced all passive relatives instead of ORs, thus showing the same pattern as her typically developing age peers. For this syntactic intervention a new game was developed ("the syntactic Scrabble"), which was useful to implicitly teach *wh*-movement without resorting to theoretical explanations. After the syntactic intervention, MM showed improved performance in all the tasks, reaching ceiling in all the structures analysed. Moreover, her narrative skills improved as well.

Before the syntactic intervention, asymmetries between production and comprehension were observed. Indeed, before treatment LB and ES did not show the same level of accuracy in the production and the comprehension tasks. After the syntactic intervention, results showed that linguistic intervention helps to eliminate the difference between the production and the comprehension of relative clauses, and both participants showed ceiling performance in both tasks.

The analysis of the three case studies suggest that children with CI can benefit from a syntactic intervention based on the explicit teaching of syntactic rules.

Indeed, they can improve their abilities in the production and comprehension of restrictive relative clauses, as is the case for LB and ES, and they can show improvement also in their narrative skills, as did ES and MM. Moreover, after treatment MM showed an increased number of subordinate clauses. This phenomenon demonstrated that her speech became more complex than before treatment.

These studies further show that a short-term intervention appears to have the same results as a long-term intervention. As pointed out above, the syntactic intervention described in this paper comprised six-to-seven sessions and lasted approximately three months, while the treatments given to patients with agrammatism (Thompson et al., 1995) and to a child with syntactic DLD (Levy & Friedmann, 2009) lasted more than six months and comprised from 16 to 42 sessions. This could be interpreted as an advantage from the clinical point of view: a short-term therapy would be more adaptable to speech therapy sessions.

Summing up, all three participants showed improvement in the production and comprehension of relative clauses after a short-term syntactic intervention. Moreover, considering that ES and MM were trained only on ORs, the results of the syntactic intervention showed generalization effects to untrained structures derived by *wh*-movement with a lesser degree of complexity than ORs (SRs), and also with a higher degree of complexity than ORs (ORps). In addition to this, ES and MM showed generalization affects also to narrative skills, such as a decrease in the production of ungrammatical and incomplete sentences, and in MM, also an increased number of subordinate clauses, which was already observed in the experiments carried out by Thompson and Shapiro (2005) on patients with agrammatism.

In conclusion, children with CIs may benefit from a treatment based on explicit teaching of syntactic rules, with improvements that can be maintained over time.

Acknowledgements

We thank the participants and their families for the collaboration, without their precious collaboration this work would not have been possible. We would like to thank the whole staff of the Ear Nose Throat Clinic, Department of Neurosciences, University of Padua.

References

Adani, F. (2011). Rethinking the acquisition of relative clauses in Italian: Towards a grammatically based account. *Journal of Child Language*, 38(1), 141–165.
https://doi.org/10.1017/S0305000909990250

Arosio, F., Adani, F., & Guasti, M. T. (2009). Grammatical features in the comprehension of Italian relative clauses by children. In J. M. Brucart, A. Gavarró, J. Solà (Eds.), *Merging features: computation, interpretation, and acquisition* (pp. 138–155). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199553266.003.0008

Ballard, K. J., & Thompson, C. K. (1999). Treatment and generalization of complex sentence production in agrammatism. *Journal of Speech, Language, and Hearing Research*, 42(3), 690–707. https://doi.org/10.1044/jslhr.4203.690

Belletti, A., & Contemori, C. (2010). Intervention and attraction. On the production of subject and object relatives by Italian (young) children and adults. In J. Costa, A. Castro, M. Lobo, & F. Pratas (Eds.), *Language acquisition and development. Proceedings of Gala 2009* (pp. 39–52). Newcastle upon Tyne: Cambridge Scholars.

Bianchi, V. (1999). *Consequences of antisymmetry: Headed relative clauses*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110803372

Carpenedo, C. (2011). *On the production of relative clauses by middle school age adolescents: An elicitation experiment* (Unpublished doctoral dissertation). Ca'Foscari University of Venice.

Caselli, M. C., Rinaldi, P., Varuzza, C., Giuliani, A., & Burdo, S. (2012). Cochlear implant in the second year of life: Lexical and grammatical outcomes. *Journal of Speech, Language, and Hearing Research*, 55(2), 382–394. https://doi.org/10.1044/1092-4388(2011/10-0248)

Chapman, R. S., & Miller, J. F. (1975). Word order in early two- and three-word utterances: Does production precede comprehension? *Journal of Speech and Hearing Research*, 18, 346–354. https://doi.org/10.1044/jshr.1802.355

Chesi, C. (2006). *Il linguaggio verbale non-standard dei bambini sordi* [*The non-standard language in children with hearing loss*]. Rome: GAIA srl-Edizioni Univ. Romane.

Chomsky, N. (1977). On wh-movement. In P. W. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal syntax* (pp. 71–132). New York NY: Academic Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Cinque, G. (1988). La frase relativa [The relative clause]. In L. Renzi, G. Salvi, & A. Cardinaletti (Eds.), *Grande grammatica italiana di consultazione: I. La frase. I sintagmi nominale e preposizionale* [*Italian big grammar for consulation: I. Sentence. Nominal Phrases and Prepositional Phrases*]. Bologna: Il Mulino.

Contemori, C., & Garraffa, M. (2010). Comparison of modalities in SLI syntax: A study on the comprehension and production of non-canonical sentences. *Lingua*, 120(8), 1940–1955. https://doi.org/10.1016/j.lingua.2010.02.011

D'Amico, S., Devescovi, A., Marano, A., & Albano, S. (2008). La valutazione della competenza narrativa in bambini prescolari e scolari attraverso un libro illustrato [The assessment of narrative skills in pre-scholar and scholar children resorting to an illustrated book]. *Rivista di Psicolinguistica Applicata*, 8(1/2), 1000–1018.

De Nichilo, A. (2017). *Insegnamento esplicito delle strutture sintattiche a movimento: Pronomi clitici, frasi passive e frasi relative in uno studente bengalese con italiano L2* [*Explicit teaching of syntactic structures derived by movement: Clitic pronouns, passivee sentences, and relative clauses in a Bengali student with Italiana s L2*] (Unpublished MA dissertation). Ca' Foscari University of Venice.

De Villiers, P. A. (1988). Assessing English syntax in hearing-impaired children: Eliciting production in pragmatically-motivated situations. *Journal of the Academy of Rehabilitative Audiology*, 21(mono.suppl.), 41–71.

Delage, H., & Durrleman, S. (2018). Developmental dyslexia and specific language impairment: Distinct syntactic profiles? *Clinical Linguistics & Phonetics*, 32(8), 758–785. https://doi.org/10.1080/02699206.2018.1437222

D'Ortenzio, S. (2018). Training the comprehension and production of restrictive relative clauses. *Generative Grammar@ Geneva*, 11,135–156.

D'Ortenzio, S., Montino, S., Martini, A., & Volpato, F. (2017). Il trattamento delle frasi relative in un bambino sordo portatore di impianto cocleare [Treatment of relative clauses in a boy with hearing impairment fitted with CI]. In F. M. Dovetto (Ed.), *Lingua e patologia. Le frontiere interdisciplinari del linguaggio* [*Language and pathology. The interdisciplinary frontiers of language*] (pp. 273–285). Rome: Aracne.

D'Ortenzio, S., & Volpato, F. (2020). How do Italian-speaking children handle wh-questions? A comparison between children with hearing loss and children with normal hearing. *Clinical Linguistics & Phonetics*, 34(4), 407–429. https://doi.org/10.1080/02699206.2019.1677779

Ebbels, S. H. (2007). Teaching grammar to school-aged children with specific language impairment using shape coding. *Child Language Teaching and Therapy*, 23(1), 67–93. https://doi.org/10.1191/0265659007072143

Ebbels, S. H. (2014). Effectiveness of intervention for grammar in school-aged children with primary language impairments: A review of the evidence. *Child Language Teaching and Therapy*, 30(1), 7–40. https://doi.org/10.1177/0265659013512321

Ebbels, S. H. (2017). Intervention research: Appraising study designs, interpreting findings and creating research in clinical practice. *International Journal of Speech-language Pathology*, 19(3), 218–231. https://doi.org/10.1080/17549507.2016.1276215

Ebbels, S. H., Van der Lely, H. K., & Dockrell, J. E. (2007). Intervention for verb argument structure in children with persistent SLI: A randomized control trial. *Journal of Speech, Language, and Hearing Research*, 50(5), 1330–1349. https://doi.org/10.1044/1092-4388(2007/093)

Ellis, R. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Bristol: Multilingual Matters.

Friedmann, N., & Novogrodsky R. (2007). Is the movement deficit in syntactic SLI related to traces or to thematic role transfer? *Brain & Language*, 101, 50–63. https://doi.org/10.1016/j.bandl.2006.09.006

Friedmann, N., Novogrodsky, R., Szterman, R., & Preminger, O. (2008). Resumptive pronouns as last resort when movement is impaired: Relative clauses in hearing impairment. *Current Issues in Generative Hebrew Linguistics*, 134, 276–290.

Friedmann, N., & Szterman, R. (2006). Syntactic movement in orally-trained children with hearing impairment. *Journal of Deaf Studies and Deaf Education*, 11, 56–75. https://doi.org/10.1093/deafed/enj002

Friedmann, N., & Szterman, R. (2011). The comprehension and production of wh-questions in deaf and hard-of-hearing children. *Journal of Deaf Studies and Deaf Education*, 16, 212–235. https://doi.org/10.1093/deafed/enq052

Garraffa, M., & Grillo, N. (2008). Canonicity effects as grammatical phenomena. *Journal of neurolinguistics*, 21(2), 177–197. https://doi.org/10.1016/j.jneuroling.2007.09.001

Gennari, S. P., & MacDonald, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, 111(1), 1–23. https://doi.org/10.1016/j.cognition.2008.12.006

Grillo, N. (2008). *Generalized minimality: Syntactic underspecification in Broca's aphasia*. Utrecht: LOT.

Guasti, M. T. (2017). *Language acquisition: The growth of grammar*. Cambridge, MA: The MIT press.
Guasti, M. T., & Cardinaletti, A. (2003). Relative clauses formation in Romance child's production. *Probus*, 15(1), 47–89. https://doi.org/10.1515/prbs.2003.005
Guasti, M. T., Papagno, C., Vernice, M., Cecchetto, C., Giuliani, A., & Burdo, S. (2014). The effect of language structure on linguistic strengths and weaknesses in children with cochlear implants: Evidence from Italian. *Applied Psycholinguistics*, 35(4), 739–764. https://doi.org/10.1017/S0142716412000562
Guida, M., Fagnani, U., & Ambrosetti, U. (2014). Impianto colceare e nucleare [Cochlear and nuclear implants]. In U. Ambrosetti, L. Del Bo, & F. Di Berardino (Eds.), *Audiologia protesica* [*Prosthetic Audiology*]. Torino: Edizioni Minerva medica.
Håkansson, G., & Hansson, K. (2000). Comprehension and production of relative clauses: A comparison between Swedish impaired and unimpaired children. *Journal of Child Language*, 27, 313–333. https://doi.org/10.1017/S0305000900004128
Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, 120(8), 1887–1897. https://doi.org/10.1016/j.lingua.2010.02.002
Jacobs, B. J., & Thompson, C. K. (2000). Cross-modal generalization effects of training noncanonical sentence comprehension and production in agrammatic aphasia. *Journal of Speech, Language, and Hearing Research*, 43(1), 5–20. https://doi.org/10.1044/jslhr.4301.05
Kayne, R. S. (1994). *The antisymmetry of syntax*. Cambridge, MA: The MIT Press.
Levy, H., & Friedmann, N. (2009). Treatment of syntactic movement in syntactic SLI: A case study. *First Language*, 29(1), 15–49. https://doi.org/10.1177/0142723708097815
Manetti, C., & Belletti, A. (2013). Causatives and the acquisition of the Italian passive. In C. Hamann & E. Ruigendijk (Eds.), *Language acquisition and development. Proceedings of GALA 2013* (pp. 282–298). Newcastle upon Tyne: Cambridge Scholars.
Martini, A., Bovo, R., Trevisi, P., Forli, F., & Berrettini, S. (2013). L'impianto cocleare nel bambino: Razionale, indicazioni, costo/efficacia [Cochlear implant in the child: Rational, incations, costs/benefits]. *Minerva Pediatrica*, 65(3), 325–339.
Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2009). Pronoun co-referencing errors: Challenges for generativist and usage-based accounts. *Cognitive Linguistics*, 20, 599–626. https://doi.org/10.1515/COGL.2009.026
Mayer, M. (1969). *Frog, where are you? New York*, NY: Dial Press.
McClellan, J., Yewchuk, C., & Holdgrafer, G. (1986). Comprehension and production of word order by 2-year-old children. *Journal of Psycholinguistic Research*, 15, 97–116. https://doi.org/10.1007/BF01067516
Penke, M., & Wimmer, E. (2018). Deficits in comprehending wh-questions in children with hearing loss–the contribution of phonological short-term memory and syntactic complexity. *Clinical Linguistics & Phonetics*, 32(3), 267–284. https://doi.org/10.1080/02699206.2017.1350883
Piccoli, E. (2018). *Ripetizione e produzione elicitata di frasi complesse in studenti adolescenti con DSA e stranieri.* Un protocollo di insegnamento esplicito [Repetition and elicited production of complex sentences in adolescent students with Learninf disorders. A protocol for the explicit teaching] (Unpublished MA dissertation). Ca' Foscari University of Venice.

Pivi, M., Del Puppo, G., & Cardinaletti A. (2016) The elicited oral production of Italian restrictive relative clauses and cleft sentences in typically developing children and children with developmental dyslexia. In P. Guijarro-Fuentes, M. Juan-Garau, & P. Larrañaga (Eds.) *Acquisition of Romance languages* (pp. 231–261). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781614513575-011

Rinaldi, P., & Caselli, M. C. (2009). Lexical and grammatical abilities in deaf Italian preschoolers: The role of duration of formal language experience. *Journal of Deaf Studies and Deaf Education*, 14, 63–75. https://doi.org/10.1093/deafed/enn019

Roth, F. P. (1984), Accelerating language learning in young children. *Journal of Child Language*, 11(1):89–107. https://doi.org/10.1017/S0305000900005602

Spenader, J., Smits, E. J., & Hendriks, P. (2009). Coherent discourse solves the pronoun interpretation problem. *Journal of Child Language*, 36, 23–52. https://doi.org/10.1017/S0305000908008854

Szterman, R., Friedmann, N. (2015). Insights into the syntactic deficit of children with hearing impairment from a sentence repetition task. In C. Hamann & E. Ruigendijk (Eds.), *Language acquisition and development. Proceedings of GALA 2013*. Newcastle upon Tyne: Cambridge Scholars.

Thompson, C. K. (2003). Unaccusative verb production in agrammatic aphasia: The argument structure complexity hypothesis. *Journal of Neurolinguistics*, 16, 151–157. https://doi.org/10.1016/S0911-6044(02)00014-3

Thompson, C. K., & Shapiro, L. P. (1995). Training sentence production in agrammatism: Implications for normal and disordered language. *Brain and Language*, 50(2), 201–224. https://doi.org/10.1006/brln.1995.1045

Thompson, C. K., & Shapiro, L. P. (2005). Treating agrammatic aphasia within a linguistic framework: Treatment of Underlying Forms. *Aphasiology*, 19(10–11):1021–1036. https://doi.org/10.1080/02687030544000227

Thompson, C. K., & Shapiro, L. P. (2007). Complexity in treatment of syntactic deficits. *American Journal of Speech-Language Pathology*, 16(1), 30–42. https://doi.org/10.1044/1058-0360(2007/005)

Vincenti, V., Bacciu, A., Guida, M., Mara, F., Bertoldi, B., Bacciu, S., & Pasanisi, E. (2014). Pediatric cochlear implantation: An update. *Italian Journal of Pediatrics*, 40. 72. https://doi.org/10.1186/s13052-014-0072-8

Volpato, F. (2010). *The acquisition of relative clauses and phi-features: Evidence from hearing and hearing-impaired populations* (Unpublished doctoral dissertation). Ca' Foscari University of Venice.

Volpato, F. (2012). The comprehension of relative clauses by hearing and hearing-impaired, cochlear-implanted children: The role of marked number features. In *Selected Proceedings of the Romance Turn IV Workshop on the Acquisition of Romance Languages* (pp. 306–325). Newcastle upon Tyne: Cambridge Scholars.

Volpato, F., & Adani, F. (2009). The subject/object relative clause asymmetry in Italian hearing-impaired children: Evidence from a comprehension task. *Studies in Linguistics*, 3, 269–281.

Volpato, F., & Bozzolan, G. (2017). Explicit teaching of syntactic movement in passive sentences and relative clauses: The case of a Romanian/ Italian sequential bilingual child. *Annali di Ca' Foscari - Serie Occidentale*, 51, 357–381.

Volpato, F., & Vernice, M. (2014). The production of relative clauses by Italian cochlear-implanted and hearing children. *Lingua*, 139, 39–67. https://doi.org/10.1016/j.lingua.2013.10.010

World Health Organization. (2018). *Deafness and hearing loss*. Geneva: World Health Organization. Retrieved from <http://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss> (3 May, 2020).

Language impairment in an Italian child with Trisomy X

Assessment and intervention

Natascha Lantschner and Anna Cardinaletti
Ca' Foscari University of Venice

This study deals with the case of Emma, an Italian 8-year-old girl affected by a sex chromosome anomaly called Trisomy X and diagnosed with Expressive Language Impairment. Language assessment of both comprehension and production was conducted through standardized and non-standardized, experimental tests. Passive, cleft, wh-sentences, and relative clauses with manipulation of number features were tested, in addition to clitic pronouns. Her spontaneous production was also assessed. Emma has problems with functional words realizing the highest portion of nominal structure: articles, prepositions, and clitic pronouns. Given her very good cognitive abilities, Emma was administered a tailor-made language intervention focused on functional elements, namely determiners and prepositions, aimed at enhancing her language awareness. During language intervention, use of Italian Sign Language (LIS) was crucial to boost lexical retrieval of functional prepositions and pluri-syllabic lexical words.

Keywords: Trisomy X, language assessment, Italian

1. Introduction

This study deals with the case of Emma, an Italian 8-year-old girl affected by a sex chromosome anomaly called Trisomy X and diagnosed with Expressive Language Impairment. The aim of the paper is twofold. First, we provide a detailed picture of Emma's language competence. To our knowledge, no analysis of the language competence of Italian-speaking individuals with Trisomy X has been provided so far. For other languages, scholars mention language impairment among the possible manifestations of the syndrome, but no in-depth analysis of language profiles of individuals with Trisomy X has been carried out. Language assessment of

https://doi.org/10.1075/lald.64.10lan

both comprehension and production was conducted through standardized and non-standardized experimental tests.

Second, we report on the language intervention which was administered to Emma. Given her very good cognitive abilities, she was administered a tailor-made language intervention focused on functional elements, namely determiners and prepositions, aimed at enhancing her language awareness. During language intervention, use of Italian Sign Language (LIS) was crucial to boost lexical retrieval of functional prepositions and pluri-syllabic lexical words.

The paper is organized as follows. In Section 2, an overview of Trisomy X is provided. Section 3 presents the participant to this study and the results of previous clinical assessment. In Section 4, materials used in this study and general results of both the first administration and the second administration are provided. Results of the first administration are presented in detail in Section 5. Section 6 is devoted to language intervention. Results of the second administration are discussed in Section 7, with a focus on articles and prepositions. Section 8 contains the general discussion, and Section 9 concludes the paper.

2. State of the art on Trisomy X

2.1 The syndrome

Trisomy X is one of the Sex Chromosome Anomalies which impact on the number and nature of chromosomes within cells. It is characterized by the presence of an extra X chromosome in females as a result of nondisjunction during meiosis and is the most common female chromosomal abnormality (Tartaglia, Howell, Sutherland, Wilson, & Wilson, 2010).[1] Trisomy X was first diagnosed in 1959 by Patricia Jacobs (Jacobs et al., 1959). Jacobs gave the females affected by the syndrome the name of "Superfemales". This term was however never accepted by the clinical community. Bernard Lennox, a British pathologist and geneticist who is the principal consultant on medical terms for Oxford English Dictionary, suggested the term that is currently being used (Lennox, 1960).

Literature reports that this Sex Chromosome Trisomy appears in about one in every 1,000 female births and that the probability of giving birth to female children with Trisomy X increases proportionally as mother's age increases (Gardner,

1. Trisomy X is also known as 47,XXX, Triple X or Triplo-X syndrome. Mosaicism is also found, namely when part of the cells present a normal karyotype (46,XX) and the others the typical Trisomy X karyotype (47,XXX), and different forms of mosaicism have been reported (Tartaglia et al., 2010).

Sutherland, & Shaffer, 2011). The syndrome is not hereditary: no data give evidence for the presence of the genetic anomaly running through families (Hook, 1983, 1992; Meschede & Horst, 1997).

Trisomy X can be diagnosed with prenatal screenings (i.e. amniocentesis or chorionic villus sampling). After birth, it can be found through blood tests that examine the karyotype. In most cases, females with Trisomy X are never diagnosed, because no symptoms emerge. Tartaglia et al. (2010) report that only 10% of subjects affected by Trisomy X are actually diagnosed.

Women with Trisomy X will be influenced by the syndrome as much as the amount of cells characterised by 47,XXX will be. There is considerable variation in the manifestation of the syndrome. Some subjects are deeply affected, while others only mildly. We consider physical, cognitive, psychological and linguistic characteristics separately.

2.2 Physical characteristics

Females born with Trisomy X are taller than average (Liebezeit, Rohrer, Singer, & Doerr, 2003; Linden, Bender, Harmon, Mrazek, & Robinson, 1988), but cases whose height is <50th percentile are also reported (Ratcliffe, Masera, Pan, & McKie, 1994). As for motor skills, literature reports a delay in starting walking, decreased muscle tone, and lack of coordination (Linden et al., 1988; Salbenblatt, Meyers, Bender, Linden, & Robinson, 1989).[2]

Patwardhan et al. (2002) found that in subjects affected by Trisomy X, the volume of cerebral mass is reduced if compared to non-affected females, but the most interesting result is that the amygdala has a reduced size if compared to 46,XX females. Warwick et al. (1999) discuss the correlation between reduced volume of cerebral mass and bad scores in tests assessing cognitive functions. Lenroot et al. (2014) report that the total grey and white matter in these subjects is 8% less than average; lateral ventricles volume is 50% more. Reduced mass and volume of both white and grey matter is true for all the cerebral regions, except for the parietal area. Thinner cortex can be found in temporal lateral lobes, and this would explain problems in language and in processing auditory information, as observed in Bishop et al. (2011). In addition, thicker cortex is present in the prefrontal medial and temporal medial cortex.

2. Literature reports many other physical disturbances associated with Trisomy X for which we refer the reader to Tartaglia et al. (2010).

2.3 Cognitive and psychological characteristics

The syndrome may manifest with mental retardation or mental health disorders (Bender, Linden & Robinson, 1993; DeLisi et al., 1994; Kusumi & Prange Jr., 1973; Linden et al., 1988; Otter, Schrander-Stumpel, & Curfs, 2010; Pennington, Puck, & Robinson, 1980; Woodhouse, Holland, McLean, & Reveley, 1992).

Cognitive deficits and learning disabilities appear more frequently in 47,XXX subjects than in the typical development population or when compared to sibling controls (Bender, Puck, Salbenblatt, & Robinson, 1990). The IQ level may vary between 55 and 115 (Bender et al., 1983; Bender, Linden, & Harmon, 2001; Geschwind, Boone, Miller, & Swerdloff, 2000; Pennington et al., 1980; Pennington, Bender, Puck, Salbenblatt, & Robinson, 1982; Pennington & Smith, 1983; Robinson, Puck, Pennington, Borelli, & Hudson, 1978). Deficits more often occur in verbal IQ than in nonverbal IQ. However, many patients have cognitive difficulties in both verbal and nonverbal cognitive domains (Bender, Linden, & Robinson, 1989; Netley, 1986; Pennington et al., 1980, 1982; Ratcliffe, 1999).

Female children with Trisomy X may develop Specific Language Impairment (SLI), but not disorders linked with autism (Bishop et al., 2011). Higher rates of attention deficits, mood disorders (anxiety and depression), and other problems related to the social cognitive sphere are also reported (Bender, Harmon, Linden, & Robinson, 1995; Lenroot et al., 2014; Linden et al., 1988).

2.4 Linguistic characteristics

Severe expressive language disorders, but preserved linguistic comprehension are often described in the literature (Garvey & Mutton, 1973; Tartaglia et al., 2010). Pennington et al. (1980), Netley & Rovet (1982) and Linden et al. (1988) highlight a correlation between subjects who present language impairments and subjects who have started walking and talking late. Bender et al. (1983) point out auditory perception problems, together with language impairment that affects both expression and reception (see also Bishop, 1987). Language impairment can continue throughout childhood into adulthood, manifesting through difficulties in language processing, verbal fluency, language comprehension, and pragmatic language, as reported in Pennington et al., (1980), Bender et al., (1983), Linden et al., (1988), and Otter et al., (2010).

Simpson et al. (2014) observe that the percentages of subjects with SLI and non-diagnosed chromosome anomaly are so high that they recommend early genetic screening to spot cases at risk of Specific Language Impairment.

3. The participant

The participant to this study is an Italian 8-year-old child, called Emma for convenience. Emma is affected by Trisomy X and diagnosed with severe Expressive Language Impairment (diagnosis code F80.1, 2018 ICD-10-CM), although she has a good cognitive level. She is exposed to both Italian and the Venetian dialect; her parents are native speakers of the dialect.

At the age of 3;3 (May 2012), Emma was assessed for the first time by speech therapists. She demonstrated good linguistic comprehension, adequate oral praxis, but extremely poor oral production especially in terms of phonological and lexical competence. Her vocabulary consisted in 16 words.[3] In November 2012, Emma was below the 10th percentile in the TCGB (Test for Grammatical Comprehension for Children; Chilosi & Cipriani, 1995/2006). In January 2013, her performance in the PPVT Peabody Test (Stella, Pizzoli, & Tressoldi, 2000) was at -1 SD. The results of the articulatory examination (Fanzago, 1983) describe a poor competence in this domain with 6% of words correctly produced. In May 2013, the speech therapist reports an IQ level of 126 in the Leiter-R (Roid & Miller, 1997), which means that her cognitive abilities are preserved. In June 2013, Emma was between the 10th and the 25th percentile in the TCGB Test, performing better than in November 2012. In the same period, at the age of 4;4, she produced the first sentences. The speech therapist reports the following example: *Mi a cacao a cafa*, "me has cocoa at home" ("I have cocoa at home").

In November 2017, the neuropsychiatrist makes a diagnosis of severe Expressive language disorder. Emma has never presented problems with communicative intentions and has always activated alternative strategies to successfully communicate. She does not renounce when she finds it hard to express her meanings. In her oral production, articles and prepositions are omitted, and verbs are often not conjugated. She has a severe phonological deficit, which makes her speech often incomprehensible. She demonstrates a good comprehension of lexical items, spatial concepts, and simple sentences. However, the specialist highlights a deficit in the morphological and syntactic domains. Indeed, Emma does not have a good performance in the comprehension of inflectional morphemes and in the computation of gender and number features. A severe deficit in verbal working memory emerges in the neuropsychiatric evaluation. She has a maximum

3. Camaioni & Di Blasio (1978) report that at 12–16 months, children have a vocabulary of 50 words. Among risk factors which can be predictors of SLI, a reduced vocabulary is reported, that is, less than 20 words at 18 months and less than 50 words at 24 months (Sabbadini & Galloni, 2005).

span of three. The specialist hypothesises that this deficit could possibly be at the basis of her language impairment.

Data collected from the case sheet compiled by the team of the Child Neuropsychiatry ward in Venice highlight that Emma's language competence is in general below average when her chronological age is considered. However, the results of the tests administered and the fact that she has begun producing sentences and phrases reveal that language acquisition is in progress, even though it is delayed and slower if compared to typically developing children. Emma's articulatory capacity is also developing. This general framework allows us to expect further improvements in Emma's language and to plan a tailor-made language intervention with the aim of promoting and enhancing her communicative and linguistic competence.

4. Profiles of the tests

In this study, Emma's language competence was assessed through the administration of standardized and non-standardized experimental tests. The contribution of the former consists in collecting data which can be compared with normative data. The latter allow us to more accurately investigate Emma's competence of complex structures of Italian. The tests we used and general results are presented distinguishing the linguistic abilities examined: oral comprehension, oral production, repetition, grammaticality judgement. Working memory and auditory discrimination have also been assessed. Finally, Emma was administered the Name-sign Test (Cagnin, 2017), with the aim of checking whether she had a better performance in the retrieval of names of the oral language or name signs of LIS. This evaluation was relevant in view of language intervention, which involved the support of LIS (see Section 6).

In Table 1, the standardized tests used are reported, together with the dates of the first and the second administration. The first administration took place from March to July 2017, the second administration after intervention, from end of September to December 2017. We also summarise the results to provide an overall picture of Emma's language competence. These results will be discussed in more detail in Sections 5 and 7 below.

In Table 2, the non-standardized tests used are reported, together with the dates of the first and the second administration and the general results. These results will be discussed in more detail in Sections 5 and 7 below.

Table 1. Summary of results in standardized tests

Standardized Assessment Tests	1st administration	2nd administration
Comprehension		
TCGB (Chilosi & Cipriani, 1995/2006)	29.03.2017 Error score: 7 Linguistic age 6;0	17.10.2017 Error score: 7 Linguistic age 6;0
TROG-2 (Bishop, 2009)	15.05.2017 Accuracy: 63,75% Standard score: 55 Linguistic age 4;6	14.11.2017 Accuracy: 78,75% Standard score: 77 Linguistic age 6;4
Peabody (PPVT) (Stella et al., 2000)	3.04.2017 Raw score: 86 Standard score: 81 (between −1 and −2 SD)	17.10.2017 Raw score: 100 Standard score: 91 (between −1 and −2 SD)
TOR (Levorato & Roch, 2007)	10.04.2017 Accuracy: 70%; Standard score: 10 53rd percentile	26.10.2017 Accuracy: 85%; Standard score: 12 87th percentile
Production		
Bus story (Renfrew, 1991)	10.04.2017 Not autonomous in retelling the story; omissions of articles and prepositions	26.10.2017 Autonomous speech; increased production of articles
Non-word repetition		
Bisiacchi, Cendron, Gugliotta, Tressoldi, & Vio (2005)	10.05.2017 0% non-words repeated correctly; 53,3% non-words repeated partially	14.11.2017 20% non-words repeated correctly; 80% non-words repeated partially
Working memory		
Bisiacchi et al. (2005)	3.04.2017 Verbal memory: Forward number span <3; backward number span 2 (below average); Visuo-spatial memory: 4 (average)	26.10.2017 Verbal memory: Forward number span 3; backward number span 2 (below average); Visuo-spatial memory: 5 (above average)
Auditory discrimination		
Bisiacchi et al. (2005)	10.05.2017 Accuracy: 83,3%	14.11.2017 Accuracy: 78,3%

Table 2. Summary of results in non-standardized tests

Non-standardized Assessment Tests	1st administration	2nd administration
Comprehension		
Comprehension of passive sentences – picture selection task (Verin, 2010)	15.03.2017 Accuracy: 97%	26.09.2017 Accuracy: 100%
Comprehension of relative clauses – referent selection task (Volpato, 2010)	20/22/29.03.2017 SR: 87,5% OR: 8,4% Fillers: 100%	26/28.09.2017 SR: 20,8% OR: 30,6% Fillers: 100%
Influence of number features in the comprehension of object relative clauses (Frugarello, 2013)	10/15.05.2017 Accuracy: 45,8% Fillers: 20,8%	23.11.2017 Accuracy: 44,8% Fillers: 45,9%
Comprehension of subject and object cleft sentences	26/31.07.2017 SC: 79,1% OC: 37,5% Fillers: 100%	5.10.2017 SC: 62,5% OC: 37,5% Fillers: 100%
Production		
Elicited production of clitic pronouns (Arosio, Branchini, Barbieri, & Guasti, 2014)	3.04.2017 Accuracy: 0%; main strategy: DP substitution (83,3%)	5.10.2017 Accuracy: 0%; main strategy: DP substitution (66,6%)
Elicited production of Wh-questions (Guasti, Branchini, & Arosio, 2012)	24.05.2017 Accuracy: 8,3% (2 SWh-Q)	14.11.2017 Accuracy: 20,8% (4 SWh-Q; 1 OWh-Q)
Repetition		
Sentence repetition test (Del Puppo, Volpato, Padovani, Zavattiero, & Lusuardi, 2016)	29.05.2017 Articles and prepositions frequently omitted; good performance in passive sentences; articles produced with an accuracy of 50%	5.12.2017 More syllables repeated in general; articles produced with an accuracy of 100%
Grammaticality judgements		
Sentences with and without articles	31.07.2017 Grammatical: 100% Ungrammatical: 0%	5.10.2017 Grammatical: 100% Ungrammatical: 0%
Sentences with and without prepositions	10.07.2017 Grammatical: 89% Ungrammatical: 0%	5.10.2017 Grammatical: 94,5% Ungrammatical: 11%
Name-sign test (Cagnin, 2017)	3.04.2017 Name signs: 75%; Italian names: 37,5%	19.10.2017 Name signs: 75%; Italian names: 50%

5. Tests before treatment

5.1 Comprehension

In the first assessment run through standardized tests, Emma demonstrates a preserved narrative comprehension in the TOR test (Levorato & Roch, 2007). She was in the 53rd percentile with an accuracy of 70%; she answered 6 out of 10 inferential questions and 8 out of 10 textual questions correctly. In PPVT (Stella et al., 2000), which assesses receptive vocabulary, she scored 86; her performance was between -1 and -2 SD, that is borderline (the standard score for her age is 100). In grammatical comprehension, Emma demonstrates poor syntactic competence. Results in TCGB (Chilosi & Cipriani, 1995/2006) show that her linguistic age was 6;0 years (delay: 2 years), while considering the TROG-2 results (Bishop, D.V.M., 2009), her linguistic age was even lower, namely 4;6 years. She passed 5/20 blocks, getting a standard score of 55. Analysing the errors committed by Emma in the TCGB, the most damaged structures are those containing verbal inflection, while nominal inflection is preserved.

As for the analysis of comprehension run through non-standardized tests, Emma demonstrated to be fully competent in passive sentences. After listening to a question like the following one,

(1) *In quale foto Marco è spinto da Sara*?
"In which photo is Marco pushed by Sara?"

Emma had to choose the correct picture out of three pictures: the target picture, a picture in which theta roles were reversed ("Marco pushes Sara"), a picture in which the agent changed ("The mother pushes Marco"). Emma chose the correct picture 97% of times (33/34).

This result is expected given her chronological age (children aged 5;3–6;2 comprehend passives 89% of times, Volpato, Verin, & Cardinaletti, 2016). Furthermore, in their guidelines for the analysis of errors in TCGB, Chilosi & Cipriani (1995/2006) report that children of 8 years perform with 94% accuracy on affirmative passives. Emma's performance is 100% in TCGB passive sentences, hence higher than what is expected by Chilosi & Cipriani (1995/2006).

In the test assessing comprehension of reversible right-branching relative clauses, (Volpato, 2010), Emma had to choose the correct referent out of 4 possible referents after listening to sentences like the ones in (2). One of the pictures matched the experimental sentence, the other represented the same action with reversed theta-roles:

(2) a. *Tocca la tigre che colpisce gli elefanti* — Subject relative
"Touch the tiger which hits the elephants"

b. *Tocca gli elefanti che la tigre colpisce* Object relative
"Touch the elephants which the tiger hits"

Number features were manipulated on both DPs (i.e., the relative head and the clause-internal DP); all four possible conditions were tested. Results show the well-known asymmetry between subject and object relatives, with better results in the former (21/24, 87,5%) than in the latter (accuracy: 3/36, 8,4%). In the filler sentences, consisting in 20 non-reversible subject relatives such as *Tocca il cane che ha l'osso in bocca* "Touch the dog that has the bone in the mouth", she performed 100%. Emma did not seem to be influenced by number features in comprehending right-branching relative clauses, differently from typically developing children whose comprehension is higher in the number mismatch condition (Adani, Van der Lely, Forgiarini, & Guasti, 2010; Volpato, 2010, 2012). In subject relatives, she performed worse in the number-mismatch condition (10/12, 83,4%) than in the number-match condition (11/12, 91,7%). No difference between the two conditions was found in object relatives. Emma's most frequent error in object relatives was choosing the correct referent in the wrong scenario, namely reversal of theta roles.

We now consider the test which directly assesses the influence of number features in the comprehension of complex sentences containing right-branching object relatives (an adaptation of Frugarello's (2013) test). In this test, the four pictures associated to each experimental sentence present different combinations of singular and plural number features on the relative head DP and the DP subject of the relative clause. An example is provided in (3). The four pictures contain the following combinations: two girls - two women (the target one), two girls - one woman, one girl - two women, one girl - one woman:

(3) *Tocca le bambine che le donne salutano*
"Touch the girls that the women greet"

The most remarkable result is the extremely poor performance of Emma in filler sentences, which consist in 24 sentences containing right-branching subject relative clauses with the same manipulation of number features (accuracy: 20,8%).

(4) *Tocca le donne che mangiano le arance*
"Touch the women that eat oranges"

This low result in filler sentences suggests that the low result in the 72 experimental sentences (45,8%) is not related to syntactic complexity but to the fact that Emma struggles with number features. Lack of verbal number agreement was also observed in her spontaneous speech.

In the test assessing the comprehension of 24 subject and 24 object cleft sentences, the two DPs occurring in the sentence could match or mismatch in number, as shown in (5). All four possible conditions were tested for both subject and object clefts:

(5) a. *È LA PECORA che lava il cavallo* Subject cleft
"It is the sheep that washes the horse."
b. *È IL NONNO che i pinguini lavano* Object cleft
"It is the grandfather that the penguins wash."

The experimenter asked the child: "There are two images. Try to tell me which one is the correct one. Now, if I tell you: It is THE SHEEP that washes the horse, which image do you select?". The choice was between two pictures, the one matching the experimental sentence and the one with reversed theta-roles. In this test, Emma demonstrated an asymmetry between subject and object clefts, with better results in the former (79,1%) than in the latter (37,5%), as expected (Del Puppo, 2016; Hirsch & Wexler, 2006). Similarly to her performance on relative clauses (see above), Emma did not use number features as cues in the comprehension of object cleft sentences. In the number-match condition, results were at chance level (50%); in the number-mismatch condition, she answered correctly 16,6% of times in the singular object - plural subject condition, and 33% in the plural object - singular subject condition.

5.2 Production and repetition

The results of the first administration of the Bus story (Renfrew, 1991) show that Emma possessed an extremely poor competence in retelling. She was not able to tell the story autonomously, and the experimenter had to intervene repeatedly to help her. Many functional words were missing. Emma produced 1 definite and 7 indefinite articles out of 28 DPs produced. Only one preposition (*con treno* 'with train') was produced.

In non-standardized tests, Emma demonstrates very poor performance in the elicited production of both clitic pronouns (0%) and Wh-questions (2/24, 8,3%). The strategy which she adopted most frequently when 3rd person singular object clitic pronouns were elicited is DP substitution (10/12; 83,3%), as shown in (6). In the other 2 cases, she produced ungrammatical sentences with clitic omission (16,6%), as shown in (7). This behaviour is well-known in school-age children with SLI (Arosio et al., 2014).

(6) *Butta il libbo* Target: *Lo butta*
"She throws the book away" "She throws it away"

(7) *Lava* Target: *Lo lava*
"She washes" "She washes it"

The test also elicited 3rd person reflexive pronouns: 5/6 of pronouns were omitted by Emma, i.e., 83,3% of times, something which is not attested in school-age children with SLI, who usually produce reflexive pronouns (Arosio et al., 2014). In 1 case, a possessive pronoun was used instead (*Mette posso suoi capelli*, "She arranges her hair", instead of *Si pettina*, "She combs herself"), again a well-known strategy.

As for Wh-questions, Emma only produced 2/12 subject Wh-questions correctly (introduced by *chi* "who"); she also produced incomplete subject questions (containing *chi* "who" or *quale* "which" and the verb, with or without the object). All object questions were incorrect. Two of them contain the wh-element and a 2nd person singular verb instead of a 3rd person plural verb to avoid the postverbal subject; e.g. *Quali bambini tiri?* 'which children do you pull?' instead of *Quali bambini tira la fatina?* 'which children does the fairy pull?').

In the repetition of non-words (Bisiacchi et al., 2005), Emma's performance was extremely poor. She did not repeat any word correctly; she only repeated 8 out of 15 words partially correctly. Normative data report that at age 8, the mean score of correctly repeated words is 13,4 (Bisiacchi et al., 2005).

In the sentence repetition task,[4] Emma produced one preposition ((*in*)*sieme* 'together' instead of *con* 'with'), and a high percentage of articles was omitted (50%). Her phonological deficit prevented her from producing a great number of words, namely pluri-syllabic words. However, Emma demonstrated good competence in the repetition of the passive sentence; she also changed the auxiliary (*venire* 'come' into *essere* 'be'), as shown in (8):

(8) *Pesca è mangiata è bambina cuolla*
peach is eaten [ɛ] child school
Target: *La pesca viene mangiata dalla bambina a scuola*
"The peach is eaten by the child at school"

This confirms data collected in the comprehension of passive sentences (see Section 5.1). As for the other sentence types, she repeated the SVO sentences rather correctly; in cleft sentences two proto-morphemes for the complementizer and one proto-morpheme for the definite article were produced (see en and e, respectively, in (9)), and long distance Wh-questions were turned into simple Wh-questions, as shown in (10):

4. A reduced version of Del Puppo et al.'s (2016) test was used, consisting in 11 SVO sentences (including 1 passive sentence and 2 sentences with embedded clauses), 2 coordinate sentences, 2 sentences with left dislocation of the object and resumptive pronoun, 3 cleft sentences, and 2 long-distance Wh-questions.

(9) *PÈ il cammello en tira col coda e mucca*
is the camel that pulls with the tail the cow
Target: *È il CAMMELLO a tirare la mucca*
"It is the camel that pulls the cow"

(10) *Qualle pessonna guadda gr rasse?*
which person looks girls
Target: *Quale persona hai detto che guardano le ragazze?*
"Which person did you say that the girls are looking at?"

5.3 Grammaticality judgements

Two grammaticality judgements tests were administered to Emma. In both cases, Emma was asked to help the experimenter correct the homework done by a little boy coming from abroad and acquiring Italian. It was explained to her that the homework could contain many mistakes and she was asked to be careful and signal them by saying "Yes" or "No" after listening to each sentence read by the experimenter.

The test focusing on articles contained 22 grammatical sentences and 22 ungrammatical sentences with missing articles; one example of each type is provided in (11):

(11) a. *Il nonno legge il giornale ogni giorno*
"The grandfather reads the newspaper every day"
b. *__ *fratello di Giovanni è marinaio*
"The brother of Giovanni is sailor"

In this test, Emma always said that the sentence was grammatical. In the test focusing on prepositions, which contains 18 grammatical sentences and 18 ungrammatical sentences with omitted prepositions as shown in (12), Emma answered that the sentence was ungrammatical twice (in both cases incorrectly). The indeed grammatical examples are provided in (13):

(12) a. *Ho mangiato la pasta con il sugo*
"I have eaten pasta with the sauce"
b. **Vado a scuola ____ la mia amica Viola*
"I go to school with my friend Viola"

(13) a. *Quei guanti servono per non bagnarsi le mani*
"Those gloves are used to not wet one's hands"
b. *D'estate è bello andare in montagna*
"In the summer it is nice to go to the mountains"

These results are in line with what is observed in her spontaneous and elicited productions, which lacked functional words.

Consider now the Name-sign test (Cagnin, 2017), which investigates the ability of memorising Italian names and name signs of LIS. On a Power Point presentation, the tested subject watches 8 name signs and hears 8 Italian names matched with pictures of faces mixed all together. The experimenter plays the presentation three times, and the tested subject is asked to memorise the 16 names. At the end of the third presentation, the test begins. The same 16 names and pictures are presented again in another order. The items consisted into 8 names that were matched with the pictures correctly, and the other 8 were matched incorrectly. The tested subject is asked to say if the match is correct or not pronouncing "Yes" or "No". In this test, Emma showed better performance in the retrieval of name signs (6/8, 75%) than in the retrieval of Italian names (3/8, 37,5%). The fact that name signs are highly iconic presumably facilitated Emma to remember them.

5.4 Working memory and auditory discrimination

Emma performed below average in the tests that assessed verbal working memory. Her score was <3 in forward number span (mean score at age 8: 4,5) and 2 in backward number span (mean score at age 8: 2,8). She was approximately average in the test of visuo-spatial working memory. She scored 4, instead of 4,3, which is the mean score for her age. In the test assessing auditory discrimination, Emma answered 31 out of 37 test items correctly, with an accuracy of 83,3%. She made 6 errors: in 4 cases, she judged as identical two words which are actually different for one sound, and in 2 cases, she judged two identical words as different. Age-peers' mean score is 34 (Bisiacchi et al., 2005). These results help us understand the nature of Emma's difficulties in speech, which could be related to articulation or phonology. Given that she also demonstrated some difficulties in sound discrimination, it could be hypothesised that her deficit is connected to phonology, and not simply to articulation.

6. Language intervention

After the first analysis, the main characteristics of Emma's language emerged, and a tailor-made intervention was planned. The project, focused on articles and prepositions, had two aims: to help Emma analyse her productions meta-linguistically and to stimulate her language awareness. The intervention relied on her good cognitive abilities and visuo-spatial working memory.

Language intervention started in May and lasted up to the end of September 2017. One hour-and-a half long sessions were planned twice a week, in a quiet room of the Child Neuropsychiatry Department of the ULSS 3 Serenissima in Venice. Emma participated in 33 sessions in total (including sessions dedicated to the administration of linguistic tests); she was absent 19 times, due to health reasons.

The supports used during language intervention were oral and written language, visual materials, body movements. Italian Sign Language was also used given that Emma had good results in the Name-sign test and the test assessing visuo-spatial working memory.

6.1 The "holes" metaphor

The "holes" metaphor has been the leading metaphor used to promote Emma's awareness of functional words. It helped her understand that her language lacked functional words, that is, her language had "holes" which had to be filled with articles and prepositions. By using the written support and visual materials, Emma was trained to the use of articles first, and then prepositions. Some examples of language training are reported below.

(14) *La bambina annusa il fiore*
"The child smells the flower"

Figure 1.

(15) a. *La palla è sopra il tavolo*
"The ball is on the table"
b. *La palla è sotto il tavolo*
"The ball is under the table"

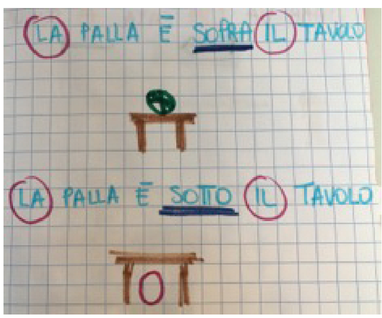

Figure 2.

The use of body movement was essential when Emma was trained to use prepositions that express movement, as in the examples below:

(16) a. *Sara va a scuola*
"Sara goes to school"
b. *Sara va da casa a scuola*
"Sara goes from home to school"

The meaningfulness of prepositions was also illustrated through minimal pairs, namely pairs of sentences which differ for only the preposition. With the support of pictures and colours (green in case of identical meaning and red in case of different meaning), their semantic contribution has been made clear. The aim was to let Emma understand that one single functional word can change the meaning of the whole sentence, as in the examples below. The sentences were accompanied by pictures: in the case of (17a), the children were together with their mum; in the case of (17b), their mother was absent:

(17) a. *I bambini fanno una torta con la mamma*
"The children make a cake with their mother"
b. *I bambini fanno una torta per la mamma*
"The children make a cake for their mother"

PPs containing functional prepositions and PPs containing adverbial prepositions were presented together associated with the same image to show that an identical meaning can be associated to different functional words:

(18) a. *In un cassetto; dentro il cassetto*
"In a drawer, inside the drawer"
b. *Sul tavolo*; *sopra il tavolo*
"On the table, on the table"

6.2 The support of Italian Sign Language

The support of LIS was crucial to elicit Emma's production of prepositions. Six signs of LIS were associated to six functional prepositions (*a* 'to', *con* 'with', *da* 'from', *da … a* 'from … to', *di* 'of', *in* 'in'). The use of signs allowed Emma to activate the lexical retrieval of functional prepositions. The same methodology was applied to lexical words to stimulate lexical retrieval in general. This revealed to be successful especially with pluri-syllabic words, which are difficult for Emma to pronounce due to her phonological deficit.

7. Results after treatment

At the end of language intervention, a second analysis was run through the same standardized and non-standardized tests used in the first administration. An improvement was detected in the results of most tests.

7.1 Comprehension

Emma's performance in narrative comprehension (TOR test) improved, reaching an accuracy of 85% (standard score 12; 87th percentile). She answered 9 out of 10 textual questions and 8 out of 10 inferential questions correctly. This result suggests that in only a few months, her language developed and improved. Emma's results in the TCGB test did not change (error score 7). However, errors were more homogeneously distributed among the structures investigated, but the main problems again emerged in verbal inflection. Accuracy in TROG-2 increased reaching a percentage of 78,75% and a standard score of 77. The number of passed blocks doubled (10/20, 50%), and the hypothesised linguistic age raised to 6;4. In the PPVT, her performance was again between -1 and -2 SD, but the standard score increased reaching 91 (close to the mean score of 100).

In the comprehension of passive sentences, results confirm that Emma is fully competent in this structure (100% accuracy). In the comprehension of reversible right-branching relative clauses, the most unexpected result is that Emma now showed a better performance in object (11/36, 30,6%) than in subject (5/24, 20,8%) relatives. As was the case in the first administration, number features did not support Emma's comprehension, although in object relatives with preverbal subjects and number mismatch, she is now more accurate (5/12, 41,6%) than in object relatives with number match (4/12, 33,3%).

In the test assessing the influence of number features in the comprehension of complex structures, the results of the first administration were confirmed. Emma still shows a problem with verbal number features, although the accuracy in the second administration of fillers slightly improved, reaching 45,9%.

As for the comprehension of cleft sentences, Emma's performance in object clefts remains the same, while her performance in subject clefts slightly worsened (from 19/24, 79,1%, to 15/24, 62,5%). These trends are in line with what was observed above for the comprehension of relative clauses. It is interesting to observe an improvement (from 16,6% to 50% of correct answers) in object clefts with number mismatch (i.e., singular focused object and plural embedded subject as in *È la giraffa che le zebre tirano*, 'it is the giraffe that the zebras pull').

7.2 Production and repetition

In the second administration of the Bus story, Emma demonstrated to be able to retell the story without any help from the experimenter. Even though her linguistic competence is extremely poor when her chronological age is considered, an improvement is evident. Emma produced more articles in general and more definite than indefinite articles (see Section 7.5 below). She also produced a functional preposition (*in* "in") and proto-morphemes (see Section 7.6 below).

In the elicited production of clitic pronouns (Arosio et al., 2014), Emma's performance confirmed the results obtained in the first administration. No clitic pronoun was produced, and the main strategy adopted was DP substitution (66,6% of times). Instead of clitic omissions, other strategies were now adopted, such as using different verbs. Reflexive pronouns were still omitted 83,3% of times. In one case, she produced a proto-morpheme for the reflexive pronoun, as shown in (19):

(19) *N taglia i capelli*
[n] cuts the hair
Target: *Si taglia i capelli*
"She cuts her hair"

The accuracy in the elicited production of Wh-questions (Guasti et al., 2012) slightly increased reaching the percentage of 20,8%. Emma produced five Wh-questions correctly, namely 4 subject questions (3 introduced by *chi* "who" and 1 by *quale*+NP "which+NP" with correct agreement) and 1 object question introduced by *chi* ("who"). In the incorrect cases, incomplete Wh-questions were produced. It is remarkable that 3rd person plural verbs were produced in this test.

In the repetition test (Bisiacchi et al., 2005), Emma repeated 3 out of 15 non-words correctly and 12 out of 15 non-words partially. A phonological improvement can be observed, even though her performance is still below average if her chronological age is considered.

An improvement from the phonological point of view can also be noticed in the sentence repetition task.[5] Emma demonstrated to repeat more syllables in general. The main result regards the production of articles. While in the first administration, she produced articles with a percentage of 50% of accuracy, in the second administration she produced them in 100% of cases (see Section 7.5 below). The two passive sentences are repeated correctly with auxiliary *venire* 'come'

5. The test was longer than in the first administration (see fn. 4): 12 SVO sentences (including 1 passive sentence and 3 containing complement clauses), 4 coordinate sentences, 5 sentences with left dislocation of the object and resumptive pronoun, 4 cleft sentences including 1 passive sentence, 7 long-distance Wh-questions, and 4 relative clauses.

being changed to *essere* 'be'. Complementizers are realized by proto-morphemes (*en, in*) and 3rd person plural verbs were also produced.

7.3 Grammaticality judgements

The results obtained in the second administration of the tests of grammaticality judgement confirmed the ones in the first administration. Emma did not seem to perceive the presence or absence of articles and prepositions. This result was not expected in the case of articles, considering that the production of these elements considerably increased (Section 7.5). Data collected in spontaneous production and sentence repetition on the one hand and in the grammaticality judgement test on the other are contrasting, namely the production of these elements precedes Emma's awareness of their presence or absence in the sentence. As for prepositions, she still incorrectly considered sentence (13a) as ungrammatical, but she now recognized 2 ungrammatical cases out of 18 (11%), slightly improving her performance.

The results of the second administration of the Name-sign test confirmed what had emerged in the first evaluation: Emma has a better performance in the retrieval of name signs (6/8, 75%) than Italian names (4/8, 50%). This result was expected, considering the good results obtained in the test assessing visuo-spatial working memory and the positive reaction she demonstrated after the introduction of LIS during language intervention.

7.4 Working memory and auditory discrimination

Data collected in the second administration of the tests assessing verbal working memory confirmed the results observed in the first administration: now visuo-spatial working memory is above average reaching a score of 5 (mean score for Emma's age: 4,3). Accuracy in the test of auditory discrimination capacities slightly decreased (78,3%). She answered 29 out of 37 items correctly, and she made 8 errors; in 100% of errors, she judged as identical two words which are actually different for one sound. Emma's performance is below average for her chronological age.

7.5 Production of articles

After language intervention, Emma's production of articles increased significantly. Furthermore, while in the first evaluation she produced more indefinite than definite articles, in the second evaluation, not only did she produce a higher quantity of articles, but she also produced more definite than indefinite articles. General results in the administered tests are reported in Table 3. The percentage of articles is calculated out of the total number of definite and indefinite DPs produced by Emma:

Table 3. Overall number of produced definite and indefinite articles

	Definite articles	Indefinite articles
1st adm.	14/52 (26,9%)	24/27 (88,8%)
2nd adm.	105/108 (97,2%)	3/3 (100%)

The increased use of articles can be observed in both spontaneous speech and the tests assessing oral production. In Table 4, the data provided in Table 3 are split into the different tasks. Note that in the second administration, Emma was asked to tell the bus story twice.

Table 4. Number of definite and indefinite articles produced in the different tests

	Bus story		Elicited production of clitic pronouns		Elicited production of Wh-questions	Sentence repetition	
	Def	Indef	Def	Indef	Def	Def	Indef
1st adm.	1/13 (7,6%)	7/9 (77,7%)	4/12 (33,3%)	11/12 (91,6%)	5/9 (55,5%)	4/18 (22,2%)	6/6 (100%)
2nd adm.	8/9 (88,8%) 5/10 (50%)	1/1 (100%) 3/3 (100%)	13/14 (92,8%)	0/0	12/13 (92,3%)	72/72 (100%)	2/2 (100%)

The data collected in the first administration of the Bus story recalls what Bottari, Cipriani, Chilosi, & Pfanner (1998: 304) report for 11 Italian children with SLI (aged 4;2–10;7, mean age 6;3). In spontaneous speech, these children produced very few definite articles (20 out of 325 definite DPs, 6.1%); and more indefinite articles were produced (21 out of 96 indefinite DPs, 21.8%). According to the authors, the semantic contribution of indefinite articles is higher than definite articles, which are fillers of the D position (Longobardi, 1994) and are thus avoided in spontaneous speech due to their scarce semantic weight. In elicitation tasks, Emma produced higher percentages of definite articles. In the second administration, the number of articles produced by Emma increased significantly, both in spontaneous speech and elicitations tasks. In the latter, she did better than the 4 Italian children with SLI (age range 7–11, mean age 9) studied by Pozzan (2007), who produced 82.4% of definite articles.

Finally consider the production of proto-articles, i.e., of proto-morphemes produced instead of the target forms, which in the case of Emma were very often vowels and nasal vowels. In the first administration of the Bus story, Emma produced 2 proto-articles out of 22 DPs (9%); in the second administration, this number raised to 2 proto-articles out of 10 DPs (20%). This means that in spontaneous speech, the production of proto-articles increased on a par with that of target articles. Two examples are provided in (20):

(20) a. *E bus va gitia*
[e] bus goes city
Target: *Il bus va in città*
"The bus goes to the city"
b. *Ha chiamato en puman*
He called [en] bus
Target: *Ha chiamato il pullman*
"He called the bus"

Data are different in the sentence repetition task. In the first administration, Emma produced 4 proto-articles out of 24 DPs (17%), in all cases instead of the definite article; in the second administration, she did not use any proto-forms anymore, confirming the attained competence; all repeated articles had the target form.

7.6 Production of prepositions

At the end of language intervention, Emma started producing functional prepositions. Earlier, she was producing proto-morphemes (e.g. *e*, *en*) most of the time, as in *Casa e legno* 'house of wood' (produced in March 2017).

In many cases, she substitutes adverbial prepositions such as *dentro* 'inside' and *sopra* 'on' for functional prepositions *in* and *su*, respectively. She also extends the substitution strategy to comitative prepositions*: insieme* 'together with' instead of *con* 'with'. Examples reported below are taken from the second administration of the Bus story (Renfrew, 1991).

(21) a. *è nato dentro tunnel*
has gone inside tunnel
"He went into the tunnel"
b. *è nato den den dentro l'acqua*
has gone in- in- inside the water (1st adm.: *va l'acqua* 'goes the water')
"He went into the water"
c. *sopra el puman*
on the bus

In the second administration of the Bus story, Emma also produces one functional preposition: *ha visso porone in puman* 'has seen owner in bus', "He saw the owner in the bus".

The most remarkable finding regards the results obtained thanks to the support of LIS. By means of signs, the lexical retrieval of prepositions has become possible for Emma. In Table 5, a transcription of Emma's spontaneous speech on October 5th, 2017 is reported:

Table 5. Spontaneous speech after intervention

Experimenter	Emma
Allora che frasi avevamo fatto l'altra volta? Hai mangiato a casa oggi o dal nonno? 'So, on which sentences did we work last time? Did you eat today at home or at grandad's?'	Casa 'home'
Rispondiamo bene 'please answer well'	A casa (without SIGN) 'at home'
Con il segno? 'with the sign?'	A casa (Emma uses the SIGN) 'at home'

The transcription in Table 6, which reports Emma's answer in the sentence repetition test, dates back to December 12th, 2017.

Table 6. Sentence repetition after intervention

Experimenter	Emma
Il papà lava la macchina rossa di mamma. "Dad washes the car red of mum"	Il papà lava la macchila di pap… la mamma 'the dad washes the car of dad … the mum'
Ok. Proviamo di nuovo con i segni. Il papà lava la macchina rossa di (SIGN) mamma "let's try again with signs. Dad washes the car red of (SIGN) mum"	Di mamma 'of mum'
Tu adesso 'you now'	Lava, lava la macchira di (SIGN) mamma 'washes, washes the car of (SIGN) mum'
Brava, bravissima! 'good, very good'	

8. Discussion

The results obtained in the tests administered to Emma suggest that the most relevant characteristics of her language competence is her phonological deficit, which prevents her from pronouncing a great number of words. This should be taken into consideration when her lexical competence is assessed. Indeed, the development of the lexicon is strictly connected to the phonological domain. D'Odorico, Bortolini, De Gasperi, & Assanelli (1999) compared typically developing children to late talkers at 24 months and demonstrated that differences in the phonological capacities are visible already at 20 months. Fasolo, Majorano, and D'Odorico (2008) confirm the influence of phonological capacities on the process of lexical acquisition. Children who are late talkers at 24 months demonstrated already at the age of 18 months phonological structures which were less complex and contained fewer consonants. These data give evidence of the correlation between Emma's poor performance in phonology and vocabulary. It must be said, however, that comparing Emma's productions at the beginning and the end of the project, a clear phonological improvement is observed. This can be seen in both her spontaneous speech and oral production tests.

As for syntactic competence, most problems are related to the functional structure of nominal expressions (DPs). At the beginning, she produced very few instances of articles and prepositions, which realize the highest functional categories of nominal expressions, and no clitic pronouns, which in many analyses also realize the highest functional categories of DPs (Cardinaletti, 2016; Torrego, 1995; Uriagereka, 1995, 2005). However, Emma did not make any nominal agreement errors in spontaneous speech or in any of the tests: Number and gender agreement is correct on articles, nouns, adjectives, demonstratives, quantifiers, and past participles. This means that the structural representation of nominal inflection is preserved. At the end of the project, articles were produced in very high percentages, and some prepositions, often in the form of proto-morphemes, are more and more produced.

As for verbal inflection, Emma's competence is delayed. Her problems with verbal inflection are not expected if her very good competence of the passive structure is taken into account. Guasti (1993/94) reports that passives are acquired by children only after the acquisition of verbal morphology. From the syntactic point of view, Emma's very good performance in passives, in both comprehension and repetition, is evidence that the structure of the inflectional phrase (IP) is acquired. We may hypothesize that her difficulties with 3rd person agreement are due to interference from the Venetian dialect, which does not distinguish between 3rd person singular and plural morphology. In the tests in which number was manipulated, namely comprehension of complex sentences containing relative and cleft clauses, Emma did not make use of number cues. Some improvement is found in the second administration as witnessed by the higher percentage of correct filler sentences in Frugarello's test and the production of 3rd person plural verbs in the test eliciting Wh-questions, the sentence repetition task, and spontaneous speech (e.g. *Ma è quelli sattano?*, 'but is those jump', "But are they those which jump?", produced in November 2017).

Moreover, observing the Wh-questions produced by Emma in spontaneous speech, as in (22a), and in the test eliciting Wh-questions, it can be said that she also possesses the complementizer phrase (CP). Some of her spontaneous productions also show that the TOP projection hosting left-dislocated elements is available in her grammar. In (22b), the left-dislocated subject *suo figlio* 'her son' precedes the Wh-word *dove* 'where', respecting the word order of the Italian left-periphery (Rizzi, 1997). Finally, sentence (22c) shows that the C position is realized by a proto-morpheme; this was also found in the sentence repetition test:

(22) a. *Quanti cavalli hai?* (20.03.2017)
"How many horses do you have?"
b. *E suo figlio dov'è?* (20.03.2017)
and her son where is?
"And where is her son?"
c. *C'è un bambino e legge* (22.03.2017)
there is a child [e] reads
"There is a child who reads"

It can be hypothesized that the syntactic structure is completely present in Emma's grammar, and that she optionally omits functional words. If the highest nodes of her syntactic tree were impaired and inaccessible, as in the Tree Pruning Hypothesis (Friedmann, 1998, 1999; Friedmann & Grodzinsky, 1997, 2000), Emma would not be able to produce Wh-questions, topicalized elements and complementizers. The phenomenon of optional omission of functional words can be explained by the Trade-off Theory (Linebarger, Schwartz, & Saffran, 1983). Due to limited linguistic resources, also considering her severe phonological disorder, Emma optionally omits functional words.

Note that the presence of proto-morphemes in Emma's productions is not expected in the light of Bottari et al. (1998: 302), who report that proto-morphemes are not attested at all in their sample of children with SLI in substitution for determiners or any other functional category. Proto-morphemes for complementizers are only found in 5% of cases in Contemori & Garraffa's (2010) study of 4 Italian children with SLI (age 4;5–5:9). The fact that Emma produces proto-morphemes for articles, prepositions, and complementizers brings her linguistic competence closer to that of (younger) children with typical development (Bottari, Cipriani, & Chilosi, 1993/94, 1996).

Consider now the decrease in the comprehension of subject relative and cleft clauses in the second administration. This trend is not new and is reminiscent of the U-shaped curve often observed in language acquisition. Levy and Friedmann (2009) report the case of Gal, a 12-year-old boy with SLI, who was exposed to language intervention focused on Hebrew complex sentences. The authors suggest that before intervention, subject relatives were processed by Gal relying on linear word order; after language intervention, the boy was influenced by the structure of object relatives and produced ungrammatical subject relatives with resumptive pronouns. A similar account can be adopted for Emma. It could be hypothesized that language intervention made her language competence develop while she first relied on linear order to understand subject relative and cleft clauses, after language intervention she started using a structural analysis to comprehend them. Since she did not rely on number features, her performance however remains low.

Note finally that Emma's performance in the sentence repetition task is higher than what is expected considering her verbal working memory (forward number span <3 in the first administration; span 3 in the second administration). This suggests that Emma's language competence allows her to process, analyse, and reconstruct the sentences that are produced by the experimenter to be repeated and confirms that repetition tasks do not simply rely on memory to be accomplished, but on language ability and in particular, on morphosyntax (cf. Marinis & Armon-Lotem, 2015; Polišenská, Chiat, & Roy, 2015, among many others).

9. Conclusion

This study has dealt with the case of Emma, an Italian 8-year-old girl affected by a sex chromosome anomaly called Trisomy X and diagnosed with Expressive Language Impairment. The systematic assessment with non-standardized test allowed us to clarify the nature of Emma's language deficit. It is remarkable that Emma comprehends and repeats passive sentences at ceiling, showing that she does not have problems with IP structure and A-movement. She also produces elements in the left-periphery of the clause, suggesting that the CP structure and A-bar movement are also in place. Her difficulties in handling verbal number inflection in the 3rd person may be attributed to interference from the Venetian dialect she speaks at home. This may have an impact on the comprehension of complex sentences where number features are manipulated. In the second administration of the tests, comprehension ameliorates, and some 3rd person plural verbal forms are produced. In addition, nominal inflection is fully preserved. Emma has problems with functional words realizing the highest portion of nominal structure: articles, prepositions, and clitic pronouns.

The 5-month-long systematic intervention aimed at enhancing language awareness led Emma become more sensitive to language structure. The main results of language intervention are twofold: the enhancement in the production of articles, and the adoption of Italian Sign Language as a strategy to boost lexical retrieval. This has been particularly useful in the case of prepositions and plurisyllabic words. Signs are included among the strategies identified to help Emma communicate. The results obtained are significant and suggest the possibility of further language improvement.

References

Adani, F., Van der Lely, H. K. J., Forgiarini, M., & Guasti M. T. (2010). Grammatical feature dissimilarities make relative clauses easier: A comprehension study with Italian children. *Lingua*, 120, 2148–2166. https://doi.org/10.1016/j.lingua.2010.03.018

Arosio, F., Branchini, C., Barbieri, L., & Guasti, M. T. (2014). Failure to produce direct object clitic pronouns as a clinical marker of SLI in school-aged Italian speaking children. *Clinical Linguistics & Phonetics*, 28(9), 639–663. https://doi.org/10.3109/02699206.2013.877081

Bender, B. G., Fry, E., Pennington, B., Puck, M., Salbenblatt, J., & Robinson, A. (1983). Speech and language development in 41 children with sex chromosome anomalies. *Pediatrics*, 71(2), 262–267.

Bender, B. G., Harmon, R. J., Linden, M. G., & Robinson, A. (1995). Psychosocial adaptation of 39 adolescents with sex chromosome abnormalities. *Pediatrics*, 96(2), 302–308.

Bender, B. G., Linden, M. G., & Harmon, R. J. (2001). Life adaptation in 35 adults with sex chromosome abnormalities. *Genetics in Medicine*, 3(3), 187–191. https://doi.org/10.1097/00125817-200105000-00007

Bender, B. G., Linden, M. G., & Robinson, A. (1989). Verbal and spatial processing efficiency in 32 children with sex chromosome abnormalities. *Pediatric Research*, 25(6), 577–579. https://doi.org/10.1203/00006450-198906000-00004

Bender, B. G., Linden, M. G., & Robinson, A. (1993). Neuropsychological impairment in 42 adolescents with sex chromosome abnormalities. *American Journal of Medical Genetics Part A*, 48(3), 169–173. https://doi.org/10.1002/ajmg.1320480312

Bender, B. G., Puck, M. H., Salbenblatt, J. A., & Robinson, A. (1990). Cognitive development of children with sex chromosome abnormalities. In C. S. Holmes (Ed.), *Psychoneuroendocrinology* (pp. 138–163). New York, NY: Springer. https://doi.org/10.1007/978-1-4612-3306-0_10

Bishop, D. V. M. (1987). The causes of specific developmental language disorder ("developmental dysphasia"). *Journal of Child Psychology and Psychiatry*, 28(1), 1–8. https://doi.org/10.1111/j.1469-7610.1987.tb00646.x

Bishop, D. V. M. (2009). *TROG 2: Test for Reception of Grammar-Version 2*. Firenze: Edizioni Giunti OS.

Bishop, D. V. M., Jacobs, P. A., Lachlan, K., Wellesley, D., Barnicoat, A., Boyd, P. A., […], & Scerif, G. (2011). Autism, language and communication in children with sex chromosome trisomies. *Archives of Disease in Childhood*, 96(10): 954–959. https://doi.org/10.1136/adc.2009.179747

Bisiacchi, P. S., Cendron, M., Gugliotta, M., Tressoldi, P. E., & Vio, C. (2005). *BVN 5-11: Batteria di valutazione neuropsicologica per l'età evolutiva*. Trento: Erickson.
Bottari, P., Cipriani, P., & Chilosi, A. M. (1993/94). Protosyntactic devices in the acquisition of Italian free morphology. *Language Acquisition*, 3(4), 327–369. https://doi.org/10.1207/s15327817la0304_1
Bottari, P., Cipriani, P., & Chilosi, A. M. (1996). Root infinitives in Italian SLI children. In *BUCLD 20: Proceedings of the 20th Annual Boston University Conference on Language Development* (Vol. 1, pp. 75–86). Somerville, MA: Cascadilla Press.
Bottari, P., Cipriani, P., Chilosi, A. M., & Pfanner, L. (1998). The determiner system in a group of Italian children with SLI. *Language Acquisition*, 7(2-4), 285–315. https://doi.org/10.1207/s15327817la0702-4_7
Cagnin, L. (2017). The name-sign test (Unpublished manuscript). Ca' Foscari University of Venice.
Camaioni, L. & Di Blasio, P. (1978). *Sviluppo del linguaggio*. Bologna: Il Mulino.
Cardinaletti, A. (2016). Sui limiti dei pronomi clitici: Inventario ed estrazione. In A. Ledgeway, M. Cennamo, & G. Mensching (Eds.), *Actes du XXVIIe Congrès International de Linguistique et de Philologie Romanes (Nancy, 15–20 juillet 2013). Section 4: Syntaxe* (pp. 48–62). Nancy: ATILF.
Chilosi, A. M., & Cipriani, P. (1995/2006). *TCGB: Test di comprensione grammaticale per bambini: Manuale*. Tirrenia: Edizioni del Cerro.
Contemori, C., & Garraffa, M. (2010). Comparison of modalities in SLI syntax: A study on the comprehension and production of non-canonical sentences. *Lingua*, 120(8), 1940–1955. https://doi.org/10.1016/j.lingua.2010.02.011
Del Puppo, G. (2016). *On the acquisition of focus: Elicited production of cleft sentences and wh-questions by school-aged, Italian-speaking children* (Unpublished doctoral dissertation). Ca' Foscari University of Venice.
Del Puppo, G., Volpato, F., Padovani, R., Zavattiero, P., & Lusuardi A. (2016). Valutare la competenza sintattica di bambini con Disturbo Specifico del Linguaggio. *Poster presented at the conference CLASTA VII* (Communication & Language Acquisition Studies in Typical & Atypical Populations), Calambrone, Pisa, 29–30 April 2016.
DeLisi, L. E., Friedrich, U., Wahlstrom, J., Boccio-Smith, A., Forsman, A., Eklund, K., & Crow, T. J. (1994). Schizophrenia and sex chromosome anomalies. *Schizophrenia Bulletin*, 20(3), 495–505. https://doi.org/10.1093/schbul/20.3.495
D'Odorico, L., Bortolini, U., De Gasperi, L., & Assanelli, A. (1999). Capacità fonologiche e sviluppo lessicale: Quale relazione. *Poster presented at the Primo Congresso Nazionale della sezione di Psicologia Clinica*, Capri, 29–30 September 1999.
Fanzago, F. (1983). Test di valutazione dell'articolazione. *Quaderni di Acta Phoniatrica Latina*, 2, 80–85.
Fasolo, M., Majorano, M., & D'Odorico, L. (2008). Babbling and first words in children with slow expressive development. *Clinical Linguistics & Phonetics*, 22(2), 83–94. https://doi.org/10.1080/02699200701600015
Friedmann, N. (1998). *Functional categories in agrammatic production: A cross- linguistic study* (Unpublished doctoral dissertation). Tel Aviv University.
Friedmann, N. (1999). "That" and "what" in agrammatic aphasia. *Brain and Language*, 69(3), 365–367.

Friedmann, N., & Grodzinsky, Y. (1997). Tense and agreement in agrammatic production: Pruning the syntactic tree. *Brain and Language*, 56(3), 397–425. https://doi.org/10.1006/brln.1997.1795

Friedmann, N. A., & Grodzinsky, Y. (2000). Split inflection in neurolinguistics. In M.-A. Friedemann & L. Rizzi (Eds.), *The acquisition of syntax* (pp. 84–104). London: Longman.

Frugarello, P. (2013). *The role of marked number features in the agrammatic comprehension of object relative clauses* (Unpublished MA thesis). Ca' Foscari University of Venice.

Gardner, R. M., Sutherland, G. R., & Shaffer, L. G. (2011). *Chromosome abnormalities and genetic counseling*. Oxford: Oxford University Press. https://doi.org/10.1093/med/9780195375336.001.0001

Garvey, M., & Mutton, D. E. (1973). Sex chromosome aberrations and speech development. *Archives of Disease in Childhood*, 48(12), 937. https://doi.org/10.1136/adc.48.12.937

Geschwind, D. H., Boone, K. B., Miller, B. L., & Swerdloff, R. S. (2000). Neurobehavioral phenotype of Klinefelter syndrome. *Developmental Disabilities Research Reviews*, 6(2), 107–116. https://doi.org/10.1002/1098-2779(2000)6:2<107::AID-MRDD4>3.0.CO;2-2

Guasti, M. T. (1993/94). Verb syntax in Italian child grammar: Finite and nonfinite verbs. *Language Acquisition*, 3(1), 1–40. https://doi.org/10.1207/s15327817la0301_1

Guasti, M. T., Branchini, C., & Arosio, F. (2012). Interference in the production of Italian subject and object *wh*-questions. *Applied Psycholinguistics*, 33(1), 185–223. https://doi.org/10.1017/S0142716411000324

Hirsch, C., & Wexler, K. (2006). Children's passives and their resulting interpretation. *The proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition–North America. University of Connecticut Occasional Papers in Linguistics*, 4, 125–136.

Hook, E. B. (1983). Chromosome abnormalities and spontaneous fetal death following amniocentesis: Further data and associations with maternal age. *American Journal of Human Genetics*, 35(1), 110–116.

Hook, E. B. (1992). Chromosome abnormalities: Prevalence, risks and recurrence. In D. J. H. Brock, C. H. Rodeck, & M. A. Ferguson Smith (Eds.), *Prenatal diagnosis and screening* (pp. 351–392). Edinburgh: Churchill Livingstone.

Jacobs, P., Baikie, A. G., Brown, W. C., Macgregor, T. N., Maclean, N., & Harnden, D. G. (1959). Evidence for the existence of the human "super female". *The Lancet*, 274(7100), 423–425. https://doi.org/10.1016/S0140-6736(59)90415-5

Kusumi, Y., & Prange Jr, A. J. (1973). Triple mosaicism of X-chromosomes with bipolar affective psychosis. *Diseases of the Nervous System*, 34(2), 94–97.

Lennox, B. (1960). Use of the term 'superfemale'. *The Lancet*, 275(7114), 55. https://doi.org/10.1016/S0140-6736(60)92744-6

Lenroot, R. K., Blumenthal, J. D., Wallace, G. L., Clasen, L. S., Lee, N. R., & Giedd, J. N. (2014). A case-control study of brain structure and behavioral characteristics in 47, XXX syndrome. *Genes, Brain and Behavior*, 13(8), 841–849. https://doi.org/10.1111/gbb.12180

Levorato, M. C., & Roch, M. (2007). *TOR. Test di comprensione del testo orale 3-8 anni*. Firenze: Giunti OS.

Levy, H., & Friedmann, N. (2009). Treatment of syntactic movement in syntactic SLI: A case study. *First Language*, 29(1), 15–49. https://doi.org/10.1177/0142723708097815

Liebezeit, B. U., Rohrer, T. R., Singer, H., & Doerr, H. G. (2003). Tall stature as presenting symptom in a girl with triple X syndrome. *Journal of Pediatric Endocrinology and Metabolism*, 16(2), 233–236. https://doi.org/10.1515/JPEM.2003.16.2.233

Linden, M. G., Bender, B. G., Harmon, R. J., Mrazek, D. A., & Robinson, A. (1988). 47, *XXX: What is the prognosis? Pediatrics*, 82(4), 619–630.
Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, 13(3), 361–392. https://doi.org/10.1016/0010-0277(83)90015-X
Longobardi, G. (1994). Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic Inquiry*, 25(4), 609–665.
Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 95–122). Bristol: Multilingual Matters.
Meschede, D., & Horst, J. (1997). Sex chromosomal anomalies in pregnancies conceived through intracytoplasmic sperm injection: A case for genetic counselling. *Human Reproduction*, 12(6), 1125–1127. https://doi.org/10.1093/humrep/12.6.1125
Netley, C. T. (1986). Summary overview of behavioural development in individuals with neonatally identified X and Y aneuploidy. *Birth Defects Original Article Series*, 22(3), 293–306.
Netley, C. T., & Rovet, J. F. (1982). Verbal deficits in children with 47,XXY and 47,XXX karyotypes: A descriptive and experimental study. *Brain and Language*, 17(1), 58–72. https://doi.org/10.1016/0093-934X(82)90005-0
Otter, M., Schrander-Stumpel, C. T., & Curfs, L. M. (2010). Triple X syndrome: A review of the literature. *European Journal of Human Genetics*, 18(3), 265–271. https://doi.org/10.1038/ejhg.2009.109
Patwardhan, A. J., Brown, W. E., Bender, B. G., Linden, M. G., Eliez, S., & Reiss, A. L. (2002). Reduced size of the amygdala in individuals with 47,XXY and 47,XXX karyotypes. *American Journal of Medical Genetics*, 114(1), 93–98. https://doi.org/10.1002/ajmg.10154
Pennington, B. F., Bender, B., Puck, M., Salbenblatt, J., & Robinson, A. (1982). Learning disabilities in children with sex chromosome anomalies. *Child Development*, 53(5), 1182–1192. https://doi.org/10.2307/1129006
Pennington, B. F., Puck, M., & Robinson, A. (1980). Language and cognitive development in 47, XXX females followed since birth. *Behavior Genetics*, 10(1), 31–41. https://doi.org/10.1007/BF01067317
Pennington, B. F., & Smith, S. D. (1983). Genetic influences on learning disabilities and speech and language disorders. *Child Development*, 54(2), 369–387. https://doi.org/10.2307/1129698
Polišenská, K., Chiat, S., & Roy, P. (2015). Sentence repetition: What does the task measure? *International Journal of Language & Communication Disorders*, 50, 106–118. https://doi.org/10.1111/1460-6984.12126
Pozzan, L. (2007). A dissociation between clitics and determiners in a group of Italian SLI children. In *Proceedings of the 8th CUNY-SUNY-NYU Mini-conference, LIBA* (Linguistics in the Big Apple: CUNY/NYU Working Papers in Linguistics). New York, NY: New York University.
Ratcliffe, S. (1999). Long term outcome in children of sex chromosome abnormalities. *Archives of Disease in Childhood*, 80(2), 192–195. https://doi.org/10.1136/adc.80.2.192
Ratcliffe, S. G., Masera, N., Pan, H., & McKie, M. (1994). Head circumference and IQ of children with sex chromosome abnormalities. *Developmental Medicine & Child Neurology*, 36(6), 533–544. https://doi.org/10.1111/j.1469-8749.1994.tb11884.x
Renfrew, C. E. (1991). *The bus story: A test of continuous speech*. Oxford: C. Renfrew.

Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar* (pp. 281–337). Dordrecht: Springer. https://doi.org/10.1007/978-94-011-5420-8_7

Robinson, A., Puck, M., Pennington, B., Borelli, J., & Hudson, M. (1978). Abnormalities of the sex chromosomes: A prospective study on randomly identified newborns. *Birth Defects Original Article Series*, 15(1), 203–241.

Roid, G. H., & Miller, L. J. (1997). *Leiter-r. Leiter International Performance Scale– Revised*. Wood Dale, IL: Stoelting CO.

Sabbadini, M., & Galloni, F. (2005). *La scala di valutazione cognitiva LEITER-R: Aspetti generali, interpretazione e modalità di somministrazione*. Berlin: Springer.

Salbenblatt, J. A., Meyers, D. C., Bender, B. G., Linden, M. G. & Robinson, A. (1989). Gross and fine motor development in 45, X and 47, XXX girls. *Pediatrics*, 84(4), 678–682.

Simpson, N. H., Addis, L., Brandler, W. M., Slonims, V., Clark, A., Watson, J., & Fairfax, B. P. (2014). Increased prevalence of sex chromosome aneuploidies in specific language impairment and dyslexia. *Developmental Medicine & Child Neurology*, 56(4), 346–353. https://doi.org/10.1111/dmcn.12294

Stella, G., Pizzoli, C., & Tressoldi, P. (2000). *Peabody Picture Vocabulary Test-Revised: adattamento italiano e standardizzazione* [*Peabody Picture Vocabulary Test-Revised: Italian adaptation and standardization*]. Torino: Omega Edizioni.

Tartaglia, N. R., Howell, S., Sutherland, A., Wilson, R. & Wilson, L. (2010). A review of trisomy X (47,XXX). *Orphanet Journal of Rare Diseases*, 11, 5–8.

Torrego, E. (1995). On the nature of clitic doubling. In H. Campos & P. Kemchinsky (Eds.), *Evolution and revolution in linguistic theory* (pp. 399–418). Washington, DC: Georgetown University Press.

Uriagereka, J. (1995). Aspects of the syntax of clitic placement in Western Romance. *Linguistic Inquiry*, 26(1), 79–123.

Uriagereka, J. (2005). On the syntax of doubling. In L. Heggie & F. Ordóñez (Eds.), *Clitic and affix combinations: Theoretical perspectives* (pp. 343–374). Amsterdam: John Benjamins. https://doi.org/10.1075/la.74.14uri

Verin, L. (2010). *Le frasi passive in età prescolare: Un test di comprensione e produzione e un'esperienza di lettura ad alta voce* (Unpublished MA thesis). Ca' Foscari University of Venice.

Volpato, F. (2010). *The acquisition of relative clauses and phi-features: Evidence from hearing and hearing-impaired populations* (Unpublished doctoral dissertation). Ca' Foscari University of Venice.

Volpato, F. (2012). The comprehension of relative clauses by hearing and hearing- impaired, cochlear-implanted children: the role of marked number features. In S. Ferré, P. Prévost, L. Tuller, & R. Zebib (Eds.), *Selected proceedings of the Romance Turn IV Workshop on the Acquisition of Romance Languages* (pp. 284–308). Newcastle upon Tyne: Cambridge Scholars.

Volpato, F., Verin, L. & Cardinaletti, A. (2016). The comprehension and production of verbal passives by Italian preschool-age children. *Applied Psycholinguistics*, 37(4), 901–931. https://doi.org/10.1017/S0142716415000302

Warwick, M. M., Doody, G. A., Lawrie, S. M., Kestelman, J. N., Best, J. J., & Johnstone, E. C. (1999). Volumetric magnetic resonance imaging study of the brain in subjects with sex chromosome aneuploidies. *Journal of Neurology, Neurosurgery & Psychiatry*, 66(5), 628–632. https://doi.org/10.1136/jnnp.66.5.628

Woodhouse, W. J., Holland, A. J., McLean, G., & Reveley, A. M. (1992). The association between triple X and psychosis. *The British Journal of Psychiatry*, 160(4), 554–557. https://doi.org/10.1192/bjp.160.4.554

Sentence reading in older adults with and without Mild Cognitive Impairment

The role of Working Memory and Interference Control

María Teresa Martín-Aragoneses[1,2], David del Río[2,3], Ramón López-Higes[3], José María Prados[3], Pedro Montejo[4] and María Luisa Delgado-Losada[3]
[1]National Distance Education University (UNED) / [2]Center for Biomedical Technology (CTB) / [3]Complutense University of Madrid (UCM) / [4]Center for the Prevention of Cognitive Impairment

While language comprehension tends to be well preserved in older adults, the processing and comprehension of syntactically complex sentences might be influenced by age-related changes in Working Memory (WM) and Interference Control (IC). Further, aging can be accompanied by cognitive decline caused by neurological conditions such as Mild Cognitive Impairment (MCI), but its impact on on-line sentence processing has rarely been studied. We report a study of word-by-word reading times and comprehension of sentences with different syntactic complexity in young adults ($n = 69$) and two subgroups of older adults: healthy older adults ($n = 32$) and older adults with single- and multi-domain amnestic MCI ($n = 21$). The experimental protocol was based on a self-paced reading task and a variety of neuropsychological measures including Operation Span (WM) and Stroop (IC) tasks. Syntactic complexity was induced using Spanish embedded relative clauses varying subject- versus object-extraction of the antecedent noun phrase (canonical or non-canonical word order, respectively). Moreover, within non-canonical sentences, we distinguished between those that did or did not contain long-distance dependencies between the extracted object and embedded verb. All these manipulations were expected to lead to a gradual recruitment of IC and WM based on the complexity of the sentence structure. Comprehension was similar across groups, with differences explained by WM capacity. In both subgroups of older adults, the on-line processing of object extracted sentences was modulated by their available IC and WM resources, although older adults with MCI seem to recruit WM to a lesser extent. In conclusion, results suggest that IC and WM have a modulatory role in

https://doi.org/10.1075/lald.64.11mar

the processing and comprehension of syntactically complex sentences in older adults. Moreover, older adults with MCI seem to be particularly overwhelmed by WM demands during sentence processing and comprehension.

Keywords: sentence reading, executive processes, aging

1. Introduction

Ageing has been associated with decreases in Working Memory (WM) and Interference Control (IC) (Borella, Carretti, & De Beni, 2008). Studies on language processing in older adults have suggested that age-related changes in these two key cognitive skills might have an impact on language performance (DeDe, Caplan, Kemtes, & Waters, 2004; See & Ryan, 1995). Here we focus particularly on sentence processing, as considerable controversy exists regarding whether (and how) general-purpose cognitive skills, such as WM, interact with basic linguistic processing at the sentence level (see Caplan & Waters, 2013, for historical background on this issue).

Moreover, in recent decades, there has been a growing interest in Mild Cognitive Impairment (MCI) as a prodromal phase of Alzheimer's disease (AD) and other age-related dementias (see Artero, Petersen, Touchon, & Ritchie, 2006; Winblad et al., 2004, for revised criteria). That is, MCI refers to a stage where increased cognitive difficulties are obvious but they do not yet interfere severely with everyday activities (Petersen, 2004). For example, a person might struggle more than usual to get the correct words or might ask repeatedly the same question during a conversation, but she or he will still able to follow the topic of the conversation without noticeable problems. Alternatively, a person might forget some important appointments, but may still able to cope with most daily tasks. Moreover, disproportionally low scores in neuropsychological tasks assessing memory, attention or language are found in such patients. Focusing on MCI offers an opportunity to study the cognitive patterns and disturbances that precede dementia in older adults, before its full development. Language processing difficulties are among the many cognitive disturbances seen in AD and MCI (Taler & Phillips, 2008). However, while the study of sentence comprehension has attracted interest in the early stages of AD (Bickel, Pantel, Eysenbach, & Schröder, 2000; Croot, Hodges, & Patterson, 1999; Marková, Horváthová, Králová, & Cséfalvay, 2017), relatively little research has addressed to this issue in MCI (see, however, Payne & Stine-Morrow, 2016).

Our interest in the current study was to examine how age- and MCI-related changes in WM and IC might affect sentence processing and comprehension. For this purpose, we used on-line measures of sentence processing (non-cumulative, self-paced reading task) and off-line measures of comprehension (post-sentence verification). Specifically, we contrasted the processing and comprehension of sentences varying in syntactic complexity in order to impose a spectrum of WM and IC demands on young adults, cognitively intact (healthy) older adults, and elderly people with MCI.

1.1 Working memory and sentence processing in older age

Sentence processing involves storage and retrieval of information about previously encountered material. Syntactic dependencies are often established between distant words in the sentence. A well-known example of such long-distance syntactic dependencies can be seen in example (1).

(1) The reporter that the senator attacked finally admitted the error.

This is an example of an object extracted from a Relative Clause (RC; *that the senator attacked*), itself embedded within a main clause (*the reporter finally admitted the error*). In this sentence type, a long-distance dependency among the initial Noun Phrase (NP) (*the reporter*), the relative pronoun (*that*) and the object position of the embedded clause verb (*attacked*) is established while the RC subject NP (*the senator*) intervenes. A kind of WM system is necessary to retrieve the information about the initial NP at the processing point of the embedded clause verb (Gibson, 1998; Lewis, Vasishth, & Van Dyke, 2006).

Compare (1) with the simpler syntactic structure of (2). This is an example of a subject-extracted from an embedded RC. The difference here is that dependencies are established between adjacent words across the RC boundary. The processing of sentences such as in (1) has been repeatedly demonstrated to be more complex compared to sentences as in (2) (Grodner & Gibson, 2005; Just et al., 1996; Kutas & King, 1995).

(2) The reporter that attacked the senator finally admitted the error.

Older adults, typically with lower WM resources than young adults, show increased problems with complex sentences such as (1) (DeCaro, Peelle, Grossman, & Wingfield, 2016; Stine-Morrow, Ryan, & Leonard, 2000). However, the WM resources devoted to on-line sentence processing (i.e., word recognition, establishment of syntactic dependencies and thematic role interpretation) might not be necessarily the same as those recruited by the strategic encoding and retrieval of sentence meaning for task accomplishment (i.e., answering questions or verbatim recall) (Caplan & Waters, 2013, 1999). In several studies, Caplan, Waters and colleagues have failed to find, in a systematic manner, a relationship between age-related reductions in WM capacity, as measured by standard verbal WM tasks, and an increased difficulty for on-line processing of complex sentences (Caplan, DeDe, Waters, Michaud, & Tripodis, 2011; DeDe et al., 2004). However, correlations between standard measures of WM capacity and the performance on off-line language tasks have been reported in their studies with older adults (DeDe et al., 2004). The authors' interpretation of these results is that on-line parsing depends on a very specialized kind of memory, different from the general-purpose WM system which is used for the encoding and access to sentence meaning (Caplan & Waters, 2013, 1999), the latter being what researchers usually identify as verbal WM. Counter to this proposal are the results in Kemper & Liu (2007), showing a modulatory role of WM capacity in sentence reading for younger and older adults. Thus, controversy exists on whether and how changes in WM might affect on-line language processing in older age.

Another issue related to the putative relationship between WM and sentence reading concerns how exactly WM might affect processing time. In general terms, one might expect that the higher the WM capacity, the faster the processing time while reading. However, this relationship is not necessarily so clear-cut, particularly in older age. For example, Stine-Morrow et al. (2000) found older adults to be less sensitive to syntactic complexity at critical points of the sentence (i.e., they showed a lower increase in processing times associated with sentence complexity as compared to young adults). This smaller effect of syntactic complexity on processing times was also accompanied by worse comprehension performance. Stine-Morrow and her collaborators concluded that the lower WM capacity of older adults impeded adequate processing of the complex material. Under this view, longer processing times might happen among people with larger WM capacity in order to aid comprehension. Unfortunately, Stine-Morrow and colleagues did not take individual measures of WM capacity in their study, so they could not provide correlations among WM capacity, processing times and comprehension, which might have reinforced such a view.

Caplan et al. (2011) examined this issue in young and older adults. They argued that if participants with larger WM resources used longer reading times to aid comprehension, then reading times should correlate *positively* with off-line performance. However, if reading times correlate *negatively* with performance, this might signal inefficient and longer processing by low WM participants. Moreover, they considered the possibility that other factors might influence comprehension, thus masking any possible correlation. They did not find a systematic correlation between reading times and performance, which argues against the idea that people might use longer on-line reading times to facilitate comprehension of syntactic complexity. However, there was an exception to this pattern in one specific kind of very complex sentences. In these sentences, longer reading times were positively associated with off-line task performance. As this only happened in one kind of sentence, the authors suggested that the increase in reading times in good performers possibly reflected a strategic encoding of sentence meaning used only for very complex material, and not the normal automatic processes related to on-line parsing.

Strategic adaptations used by older adults to help encode sentence meaning have been previously reported and are often reflected in end-of-clause wrap-up effects (Stine-Morrow & Payne, 2016). The wrap-up effect refers to increased processing times at the end of clause boundaries, which purportedly reflect integrative processes of clause and sentence meaning. Payne, Gao, Noh, Anderson, and Stine-Morrow (2012) found that older adults with higher WM showed longer reading times at clause and sentence wrap-up than those with lower WM, suggesting a compensatory use of WM (although this effect was lower for older adults with higher reading experience). Similarly, Nicenboim, Logačev, Gattei, and Vasishth (2016) recently found that longer dependencies produced a slowdown in reading times for young adults with higher WM capacity, but not in those with lower WM capacity.

In summary, the existing literature has examined cases where higher WM might aid on-line sentence processing. However, it has not been established with any certainty whether this might result in increased reading times (promoting compensatory recruitment of processing time and/or integrative processes of sentence meaning) or decreased reading times (making on-line computations more efficient).

1.2 Interference Control and sentence processing in older age

Older adults tend to have problems with IC; in other words, in inhibiting no-longer relevant information (Hartman & Hasher, 1991; Hasher, Lustig, & Zacks, 2007). Many studies suggest that executive skills might play an important role in language processing (Goral et al., 2011; López-Higes, Prados, Rubio, Montejo, & del Río, 2016; Yoon et al., 2015). In fact, it has been suggested that age-related differences in the efficiency of inhibition rather than in WM capacity might explain language processing difficulties in older adults. In a study by See & Ryan (1995), which explored the mediating role of WM, IC and processing speed in language performance across young and older adults, it was found that although each of these factors significantly explained the observed variation in language performance, WM did not make a significant contribution after IC and processing speed had been controlled for.

With respect to sentence comprehension, IC skills might contribute to sentence processing in multiple ways. For example, as sentence comprehension proceeds incrementally, difficulties might arise when incoming information interferes with previous information that has already been processed. Similarly to what happens in the Stroop color-naming task, where the participant has to inhibit word reading to name the color of ink (e.g., for the word *red* written in blue ink), IC might be crucial for language comprehension when the hearer/reader should abandon an incorrect interpretation to embrace a new one (Hsu & Novick, 2016; Vuong & Martin, 2014). Because of WM constraints, parsing mechanisms have been suggested to preferentially opt for simpler sentence structures. De Vincenzi (1991, 1996), for example, proposes that the parser tends to postulate dependencies with the minimal complexity (the Minimal Chain Principle), leading to a preference to assume subject extractions when an argument establishes a dependency within another clause. When incoming material shows unequivocally that the final sentence structure turns out to be more complex, that in turn triggers the necessity to reanalyze and reinterpret the sentence. Sentences such as (1) above (*The reporter that the senator attacked finally admitted his error*) are complex not only because of higher WM demands. When encountering the RC, the reader should discard the first NP (*the reporter*) as the subject and agent and embrace a new analysis where it is the object of the RC. In this regard, previous results with young adults have shown that the less suitable an initial NP is as an agent (e.g., an inanimate NP as in *The idea that the senator attacked was demonstrated to be right*), the easier it is to overcome the initial interpretation, facilitating processing (Baudiffier, Caplan, Gaonac'h, & Chesnet, 2011; Betancort, Carreiras, & Sturt, 2009; Mak, Vonk, & Schriefers, 2006).

Recent results suggest that older adults with better IC skills (Goral et al., 2011; Yoon et al., 2015) show better sentence comprehension performance. However, these studies considered the off-line completion of an acceptability task and did not study whether an on-line modulation of processing difficulty by IC skills also occurred.

1.3 Working Memory, Interference Control, and language in MCI

Older adults with MCI and people in the early stage of AD show increased WM problems in comparison to healthy older adults, and also difficulties related to IC and executive function (Aurtenetxe et al., 2016; Spieler, Balota, & Faust, 1996; Traykov et al., 2007). While sentence comprehension deficits have been found among people with AD, increasing in the course of the disease (Bickel et al., 2000; Croot et al., 1999; Marková et al., 2017), difficulties in sentence comprehension in MCI have been addressed much less.

Language comprehension deficits in AD are multifaceted (Taler & Phillips, 2008), but difficulties with complex sentences have been linked to WM problems

(Bickel et al., 2000; Small, Kemper, & Lyons, 2000). However, other studies suggest that attentional and executive problems beyond WM might also play a role for sentence comprehension difficulties in AD (Croot et al., 1999; Grossman & Rhee, 2001).

Moreover, the same controversy surrounding the use of WM resources for the on-line computation of sentence structure holds in pathological ageing. For example, using self-paced listening, Waters & Caplan (2002) found that people with AD were as sensitive to syntactic complexity during on-line sentence listening as healthy older adults, although final performance was lower for people with AD. This is in agreement with the idea of a selective sparing of on-line processing resources for parsing, and a depletion in the use of WM for the encoding and retrieval of off-line meaning.

In summary, the evidence for the role of cognitive skills in the on-line processing of sentences in MCI is scarce. Furthermore, to the best of our knowledge, no study has systematically addressed the role of IC skills and WM capacity on the on-line modulation of processing difficulty in older adults and in people with MCI.

1.4 The current study

Inspired by the outstanding research questions discussed above, the current study aims to investigate how WM and IC might influence the processing of sentences of different syntactic complexity across young and older adults, and within older adults, to explore the difficulties related to MCI. To do this, we contrasted word-by-word reading times and comprehension of three different kinds of Spanish sentences with different levels of syntactic complexity (del Río, López-Higes, & Martín-Aragoneses, 2012). Example sentences are shown in Table 1.

There are two factors affecting these syntactic structures (see the two last columns in Table 1), which might contribute to the recruitment of WM and IC. One is non-canonical word order. Canonical, unmarked word order in Spanish, involves a Subject-Verb-Object (SVO) order of major syntactic constituents, as in sentence (3a). However, under some circumstances, the object might appear before the subject, as in (3b), where $_{\text{ACC}}$ stands for accusative case.

(3) a. Juan saludó a María.
Juan greeted María.
b. La saludó Juan.
Her$_{\text{ACC}}$ greeted Juan.
Juan greeted her.

There is a strong preference to consider the first, animate NP of a sentence in Spanish as the subject and agent (Betancort et al., 2009; Casado, Martín-Loeches,

Table 1. Overview and examples of the experimental materials

Sentence type	Example	Canonical word order	Long-distance dependency
SR	El reportero que atacó al senador admitió su error. The reporter that attacked to-the$_{\text{ACC}}$ senator admitted his error. *The reporter that attacked the senator admitted his error.*	Yes	No
ORsi	El reportero al que atacó el senador admitió su error. The reporter to-the$_{\text{ACC}}$ that attacked the senator admitted his error. *The reporter whom the senator attacked admitted his error.*	No	No
ORst	El reportero al que el senador atacó admitió su error. The reporter to-the$_{\text{ACC}}$ that the senator attacked admitted his error. *The reporter whom the senator attacked admitted his error.*	No	Yes

SR: Subject Relative; ORsi: object relative with subject inversion; ORst: Object Relative with subject topicalization; ACC: accusative.

Muñoz, & Fernandez-Frías, 2005; del Río et al., 2011; del Río et al., 2012), unless syntactically marked otherwise, as in example (3b), or semantically implausible, as with initial inanimate NPs.

Spanish sentences with a subject-extracted RC (SR) such as (4a) might be considered syntactically simpler and easy to process, as each new incoming constituent is straightforwardly integrated. On the contrary, sentences containing an object-extracted RC, such as (4b) and (4c) might be considered syntactically more complex. They involve a non-canonical order of arguments, in which an initial, animate object NP appears before the subject.

(4) a. El reportero que atacó al senador admitió su error.
The reporter that attacked to-the$_{\text{ACC}}$ senator admitted his error.
The reporter that attacked the senator admitted his error.
b. El reportero al que atacó el senador admitió su error.
The reporter to-the$_{\text{ACC}}$ that attacked the senator admitted his error.
The reporter whom the senator attacked admitted his error.
c. El reportero al que el senador atacó admitió su error.
The reporter to-the$_{\text{ACC}}$ that the senator attacked admitted his error.
The reporter whom the senator attacked admitted his error.

As discussed above, in order to correctly process these two sentence types, the reader should inhibit the tendency to consider the first NP as the subject and agent of the embedded clause verb, which might require the engagement of IC skills (see Baudiffier et al., 2011; Betancort et al., 2009; or Di Domenico & Di Matteo, 2009; for similar manipulations in French, Spanish or Italian, respectively). Furthermore, given that non-canonical word order implies processing costs associated with deviations from the expected canonical pattern, these sentence structures might also require the recruitment of WM so that the reader is able to backtrack or rehearse information (Rogalsky, Matchin, & Hickok, 2008).

Another important factor is the presence of long-distance syntactic dependencies. An interesting feature in Spanish object-extracted RC is that they offer the possibility to alternate between subject-inversion (ORsi) and subject-topicalization (ORst), depending on whether the embedded subject is positioned after or before the embedded verb (Gutiérrez-Bravo, 2003), as in (4b) and (4c), respectively. Thus, long-distance dependencies occur across the RC in ORst, but not in ORsi. As discussed earlier, when encountering the embedded verb in sentences like (4c), the reader should establish a long-distance dependency, retrieving information about the extracted object with the interposed interference of the embedded subject. Therefore, it is expected that ORst, which implies interference-based, non-local integration costs, imposes higher WM and IC demands in comparison with ORsi (Gibson, 1998; Van Dyke, 2007).

As a secondary aim, we also wanted to test how on-line processing difficulties and comprehension performance were related. Therefore, we included an exploratory analysis on how accuracy on comprehension was influenced by reading times in each group. According to a compensatory view of longer processing times, trials with longer processing times will be often associated with comprehension hits, at least when compensatory recruitment of processing time is successful. These longer processing times might be used by older adults to maintain adequate performance (Stine-Morrow et al., 2000), particularly among older adults with good cognitive skills. Nonetheless, unsuccessful compensatory processing is also a possibility. In this case, readers with better skills might have longer processing times, but these would not be necessarily tied to good comprehension. Finally, an inefficient view of longer processing times assumes that people with lower cognitive skills will take more time in complex sentences and be more likely to fail. This view predicts that the longer processing times will occur in trials leading to failure, and especially in older adults with lower WM and IC skills.

2. Method

2.1 Participants

A total of 122 participants took part in this study. Sixty-nine were young adults (age range 19–29 years old; 60 women, 9 men) recruited from Complutense University of Madrid and Camilo José Cela University, who received course credit for their participation. Fifty-three were older adults (age range 60–80 years old; 36 women, 17 men) recruited at the Center for Cognitive Impairment Prevention, a Public Health Institute of the Madrid City Council. This group of older adults underwent a detailed neurological and neuropsychological assessment (for more information, see below). According to the results, they were distributed in two groups: cognitively intact (i.e., healthy) older adults ($n = 32$) and older adults with MCI ($n = 21$). Within this latter group, 12 participants were identified as single-domain amnestic MCI and 9 as multi-domain amnestic MCI by experts from Center for Cognitive Impairment Prevention. All older adults were further required to have more than 6 years of schooling and no signs of dementia. All participants were native Spanish speakers and had normal or corrected-to-normal vision, and none of them reported history of previous neuropsychiatric impairment.

2.2 Procedure

2.2.1 *Neuropsychological assessment of older adults*

Older adults underwent a detailed assessment including Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975; Spanish adaptation by Lobo et al., 1999), Global Deterioration Scale (GDS; Reisberg, Ferris, De Leon, & Crook, 1982), Yesavage Geriatric Depression Scale - Short Version (GDS-15; Sheikh & Yesavage, 1986), and the Logical Memory (LM) subtest of the Wechsler Memory Scale - Third Edition (WMS-III; Wechsler, 1997), among other questionnaires and tests used for clinical diagnosis by practitioners from the Center for Cognitive Impairment Prevention.

2.2.2 *Working Memory assessment*

To measure individual WM capacity, we used an adapted version of the Operation Span Task (OST; Unsworth, Heitz, Schrock, & Engle, 2005). The task consisted of sixty unrelated sums of two one-digit numbers and their results. In the half of the sums, one of the addends was replaced by another number to create easily detectable arithmetic errors (e.g., 3 + 5 = 9), while the other half was easily identifiable as correct (e.g., 5 + 7 = 12). Sums were randomly assigned to five different blocks, with each block consisting of three different trials. The size of trials incrementally ranged from 2 to 6 sums across the blocks. Thus, the first block begun with trials of two sums, the trial size being increased in one sum as participant progressed from one block to the following, up to a maximum span of six sums per trial.

These sums were displayed one by one in the centre of a computer screen using Superlab software (Cedrus corp., San Pedro, CA). Participants were asked to judge each sum's acceptability, and to remember the letters which were presented on the screen after each sum. Once participants gave their response to the sum by pressing a key, a capital letter was flashed in the centre of the screen for 1 second. At the end of each trial, a red question mark appeared on the computer screen, indicating to participants that they should try to remember each one of the letters displayed, in the same order as they had appeared across the trial. Participants were instructed to give their responses aloud, so that they could be recorded by the experimenter on a separate answering sheet.

Task administration was interrupted when the participant was unable to recall the whole set of letters presented in at least two of three trials of the same size. WM capacity was expressed as the total number of letters correctly recalled by the participant.

2.2.3 *Interference Control assessment*

Participants also completed the Spanish version of the Stroop Test (Golden, 2001). This is a widely used neuropsychological test to assess the ability to suppress stimulus-bound responses and resist interference (Homack & Riccio, 2004). The interference index for the Stroop test proposed by Chafetz & Matthews (2004) was used as a measure of Interference Control (IC).

2.2.4 *Self-paced reading*

Thirty triplets of sentences containing embedded RCs like those shown in Table 1 were generated. RCs were semantically reversible (i.e., either the subject or the object could plausibly assume the role of agent performing the action). Each item of a triplet was assigned to one of three blocks, so that only one version of each sentence appeared in each block. Experimental items were interspersed with 46 filler items using various different syntactic structures. Each participant was randomly assigned to one of the three blocks and the order of sentence presentation was randomized.

Sentences were presented using Linger (Rhode, 2003; <http://tedlab.mit.edu/~dr/Linger/>) according to a non-cumulative self-paced word-by-word moving window paradigm (Just, Carpenter, & Woolley, 1982). They were presented one at a time in the center of a computer screen with all characters replaced by a series of dashes, except for the blank spaces. Participants pressed the spacebar of the computer keyboard to reveal each word in the sentence. They were instructed to read the sentences for comprehension at a normal pace. Time between key-presses served to estimate the reading time per word. Before the experimental session began, participants performed a short practice block composed of 21 sentences.

2.2.5 *Off-line sentence comprehension assessment*

To evaluate reading comprehension, a verification probe appeared at the end of some sentences. Specifically, 20 experimental and 24 filler sentences carried out a question probe at the end. Reading comprehension was not assessed after each sentence in order to avoid fatigue, particularly among older adults. Each probe consisted of a single assertion about the content of the preceding sentence. Participants were instructed to press one of two different buttons on the keyboard to indicate whether the verification probe was "true" or "false". For experimental items, the half of the verification probes assessed thematic role assignment of the RC, and the other half did so with respect to the main clause. "False" probes concerning RCs involved the reversal of thematic roles (e.g., the patient acting as the agent), while "false" probes concerning the main clause used the RC-internal NP as the main subject. Feedback about the accuracy of responses was provided only during practice trials.

3. Results

3.1 Sociodemographic and cognitive differences among groups

Differences among groups in sociodemographic variables, such as age and schooling level as well as in cognitive variables, were tested with ANOVAs, and further assessed through pairwise comparisons with Bonferroni adjustments. Additional differences between both groups of older adults on MMSE, Yesavage GDS-15, and delayed recall in LM of WMS were examined by *t*-test.

Table 2 displays means and standard deviations (*SD*) for relevant sociodemographic and cognitive variables across the three groups. By design, age differed between groups [$F_{(3,118)} = 1833.3$; $p < .001$]. These differences in age were statistically significant between young adults and both healthy older adults ($p < .001$) and participants with MCI ($p < .001$), but not between both groups of older adults. Statistically significant differences were also observed in relation to years of schooling [$F_{(3,118)} = 20.83$; $p < .001$]. Specifically, both groups of older adults had similar years of schooling; but, in contrast, young adults had more years of schooling than both groups of older adults ($p < .001$, in both cases). In order to control for possible effects of years of schooling in age-related differences, analysis of reading times and sentence comprehension were carried out with years of schooling as a covariate.

Regarding cognitive differences between groups of older adults, participants with MCI showed, as expected, poorer performance on the MMSE ($p < .001$), greater difficulty recalling episodic information ($p < .001$), and also an increased depressive symptomatology, as assessed by Yesavage questionnaire ($p < .01$).

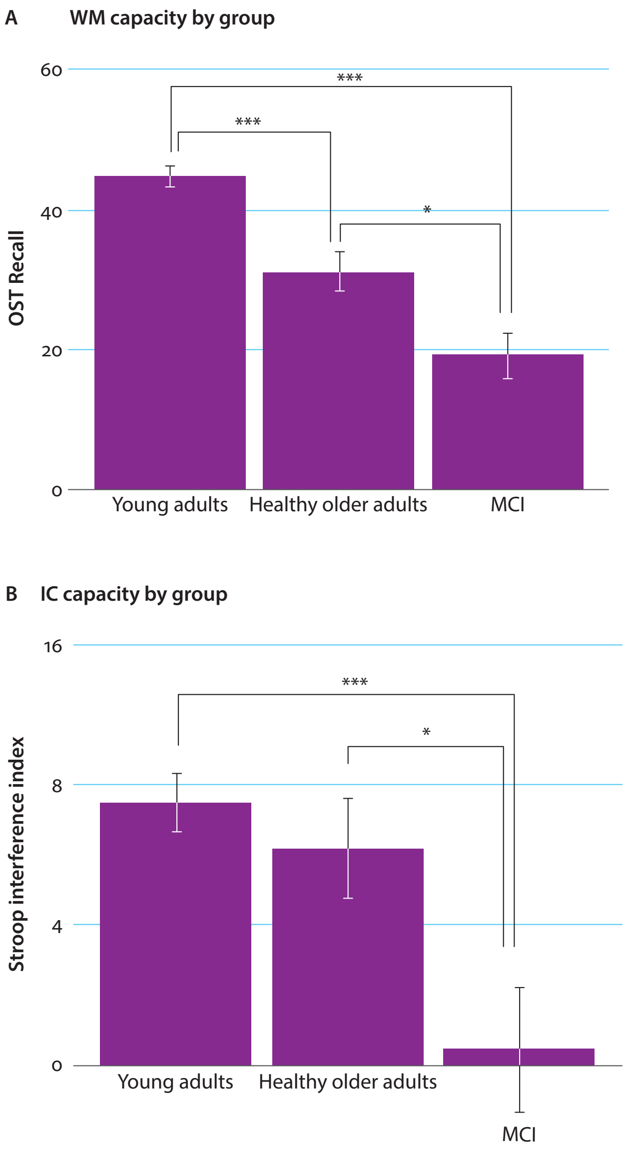

Figure 1. Mean scores in WM capacity and IC capacity across groups
Note: Error bars represent standard error of the mean. ***$p < .001$; **$p < .01$; *$p < .05$.

Table 2. Summary of sociodemographic and neurocognitive data

	Mean (±*SD*)		
	Young adults	**Healthy older adults**	**MCI**
Age (in years)	20.67 (2.47)	70.34 (4.99)	73.28 (4.87)
Schooling (in years)	14.19 (1.82)	9.87 (3.10)	10.28 (4.79)
MMSE	–	29.12 (1.10)	26.52 (2.38)
Yesavage GDS-15	–	1.34 (1.79)	3.05 (2.35)
LM delayed recall (WMS III)	–	23.81 (10.47)	5.52 (4.74)

As can be appreciated in Figure 1, there was a clear effect of group on WM capacity [$F_{(3,118)} = 23.765$; $p < .001$]. Pairwise comparison showed a stepwise effect, with healthy older adults scoring lower than young adults ($p < .001$), and in turn participants with MCI scoring lower than healthy older adults ($p < .05$). A group effect was also observed in IC [$F_{(3,118)} = 4.923$; $p < .01$]. In this case, participants with MCI scored lower than the other two groups on IC ($p < .001$ with regard to young adults, and $p < .05$ with regard to healthy older adults). However, pairwise comparisons revealed no difference in IC between young adults and healthy older adults.

3.2 Self-paced reading

Reading times below 150 ms and above 10000 ms across the whole dataset were discarded from analysis (less than 0.2% of the total data). Further, after excluding these extreme values, reading times above 3 *SD* from each individual's mean were removed. This trimming procedure affected 2.06% of the whole dataset, including experimental and filler items. Next, the effect of word length was regressed out from individual reading times (Trueswell, Tanenhaus, & Garnsey, 1994) after log transformation. So, a regression equation predicting log-transformed reading times from word length was calculated for each participant using his/her whole data set. Finally, residuals of log-transformed reading times per word were obtained by subtracting predicted from actual reading times.

Sentences were segmented into four critical regions for analysis: (a) the onset of the RC, which included the RC pronoun in subject-extracted sentences, and the preposition plus article in object-extracted sentences; (b) the initial region of the RC, which comprised the two first words of the RC (one content word and one function word across the different conditions), (c) the final word of the RC (RC closure), and (d) the main clause verb. The segmentation of sentences is shown in Table 3. Figure 2 shows residual reading times per region and sentence type across the three groups.

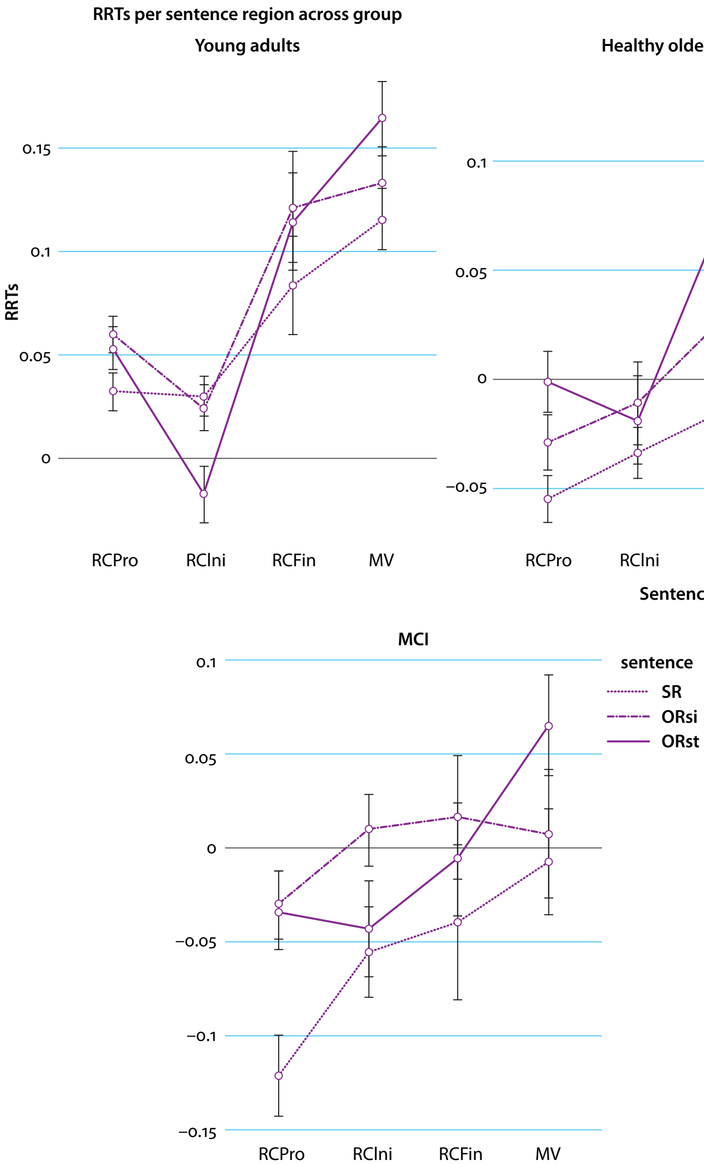

Figure 2. Mean residual reading times
Note: Error bars represent standard error of the mean.

Table 3. Regional segmentation of sentences for analysis

		RC pronoun	RC initial region	RC closure	Main verb	
SR	(El periodista) *(The journalist)*	que *that*	atacó al *attacked the*	senador *senator*	admitió *admitted*	(su error.) *(his error.)*
ORsi	(El periodista) *(The journalist)*	al (que) *whom*	atacó el *attacked the*	senador *senator*	admitió *admitted*	(su error.) *(his error.)*
ORst	(El periodista) *(The journalist)*	al (que) *whom*	el senador *the senator*	atacó *attacked*	admitió *admitted*	(su error.) *(his error.)*

Note: words in brackets were not included in the analysis.

Effects of sentence structure, WM capacity, IC skills and group on residual reading times were analyzed at each sentence region using linear mixed models (Baayen, Davidson, & Bates, 2008) by means of the statistical package lme4, available in the R programming environment (R Core Team, 2018).

Sentence structure types were Helmert-coded to assess the effect of non-canonical word order and the effect of long-distance integration within object-extracted clauses. That is, a first contrast level tested whether reading times for both complex object-extracted RCs (ORsi and ORst) were different from the simpler subject-extracted RC (SR), while a second contrast level tested whether reading times for ORst sentences were different from those for ORsi sentences. Helmert coding was also used to explore the effect of group. A first contrast tested differences between young adults and both groups of older adults (that is, a global effect of age). The second contrast tested differences between older adults depending on the presence and absence of MCI (see Table 4). WM and IC scores were used as linear continuous predictors and their values were scaled to z scores before entering into the models.

Table 4. Helmert coding for the contrasts concerning sentence type and group

Sentence type	Contrast		Group	Contrast	
	Non-canonical	Long-distance		Age	MCI
SR	–0.5	0	Young adults	–0.5	0
ORsi	0.25	–0.5	Healthy older adults	0.25	–0.5
ORst	0.25	0.5	MCI	0.25	0.5

Linear mixed effect models were constructed based on interactions among sentence structure (word order and long-distance effects), group (age and MCI effects) and cognitive skills (WM capacity and IC skills, including years of schooling as a covariate). Estimates with an absolute value of the *t*-statistic above 2 were considered statistically significant (Baayen, 2008). The maximal random effect structure leading to a convergent model was used for linear mixed effect modelling, which always involved random slopes and intercepts by subject (Barr, Levy, Scheepers, & Tily, 2013). Appendix 1 presents a summary of fixed effect results across the different sentence regions of interest. Variance inflation across all models was never above 2.75.

Given ORsi and ORst were equal in terms of integration costs up to the region of the RC pronoun, no differences due to a long-distance effect were expected at this point, and only effects of non-canonical word order were considered at the sentence level for this region. Results show an effect of non-canonical word order, as well as effects of age and years of schooling, but no interactions of interest.

Crucially, several statistically significant three-way interactions were found among sentence structure, group and cognitive skills across the RC and at the main verb. Figures 3 and 4 show these interactions.

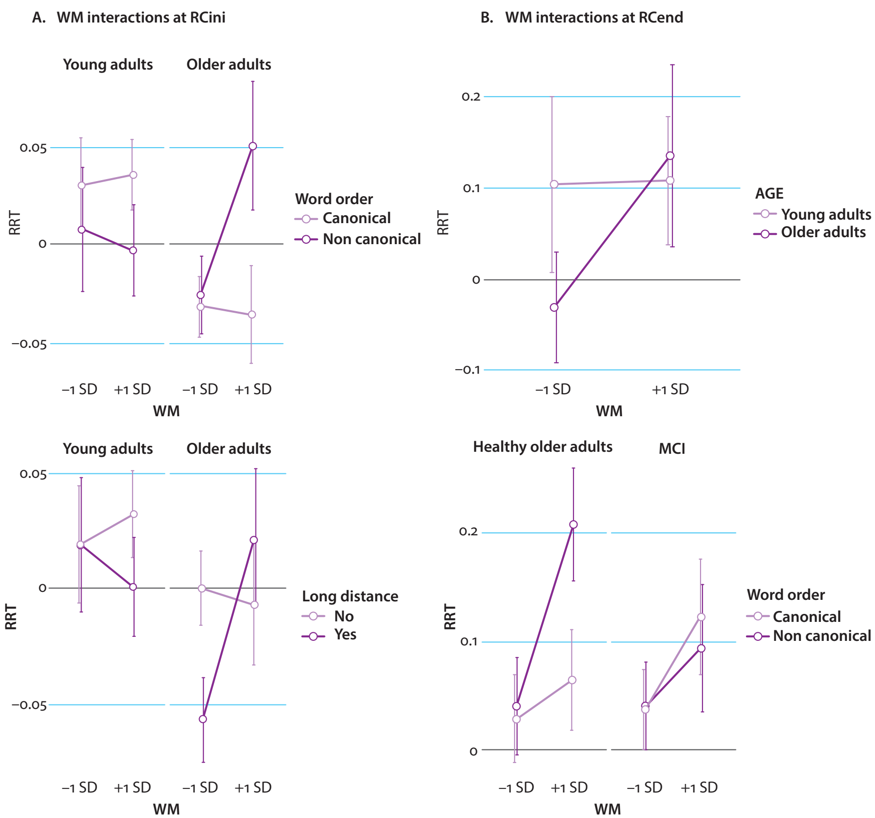

Figure 3. Interactions of WM, sentence type and group across the relative clause (RC)
Note: Effects are estimated at ±1 SD of WM scores. Error bars represent the standard error of the parameter estimates.

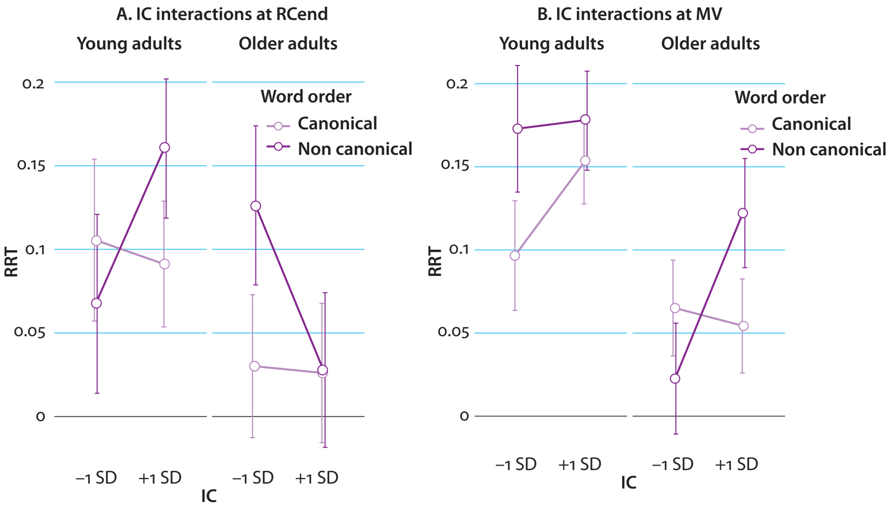

Figure 4. Interactions of IC, sentence type and group
Note: Effects are estimated at ±1 SD of IC scores. Error bars represent the standard error of the parameter estimates.

Regarding WM interactions, older adults with higher WM capacity dedicated longer residual reading times to non-canonical sentences at the initial region of the RC (Figure 3a, upper panel). This effect was more pronounced for the more complex ORst sentences containing long-distance dependencies, where older adults with low WM dedicated less processing time (Figure 3a, lower panel). At the end of the RC, older adults with low WM dropped their residual reading times in comparison to older adults with high WM capacity (Figure 3b, upper panel). This effect observed in older adults was also influenced by the presence of MCI and non-canonical word order: only healthy older adults with greater WM scores showed larger residual reading times for both types of non-canonical sentences. This was not the case for older adults with MCI (Figure 3b, lower panel).

Regarding IC, there was an interaction of IC and age with sentence type, so that older adults with low IC skills (irrespective of the presence of MCI) showed an increase in residual reading times for non-canonical sentences at the end of the RC (see Figure 4a). But this effect was reversed at the subsequent region of the main verb (see Figure 4b).

3.3 End-of-sentence verification probes

Effects of fixed factors (sentence structure, group, IC and WM) on the accuracy of post-sentence comprehension probes were tested through generalized linear mixed models for binomial outcomes (Jaeger, 2008). Again, the statistical package lme4, available in the R programming environment (R Core Team, 2018), was used for data analysis. In the same vein as in the reading time analysis, effects of sentence structure and group were coded using Helmert contrasts. So, for sentence structure, a first contrast tested differences between sentences with subject-extracted RC and object-extracted RC (i.e., effect of word order), while a second contrast assessed differences between ORsi and ORst sentences (i.e., effect of long-distance dependency). For group, a first contrast tested differences between young and older adults, regardless of the presence of MCI (i.e., effect of age). The second contrast assessed differences between the groups of older adults: healthy older adults vs older adults with MCI (i.e., effect of MCI). WM and IC were scaled to standard scores before entering into the models. Models were constructed based on interactions among sentence structure (word order and long-distance effects), group (age and MCI) and cognitive skills (WM capacity and IC skills, including years of schooling as a covariate).

Different models were estimated for question probes depending on whether thematic assignment concerned the embedded or main clause, as they were subjected to different cognitive demands. Only a random intercept was modeled by subjects and items to ensure model convergence due to a low number of items by clause type. The variance inflation factor was below a maximum of 2.98. The results are presented in Appendix 2.

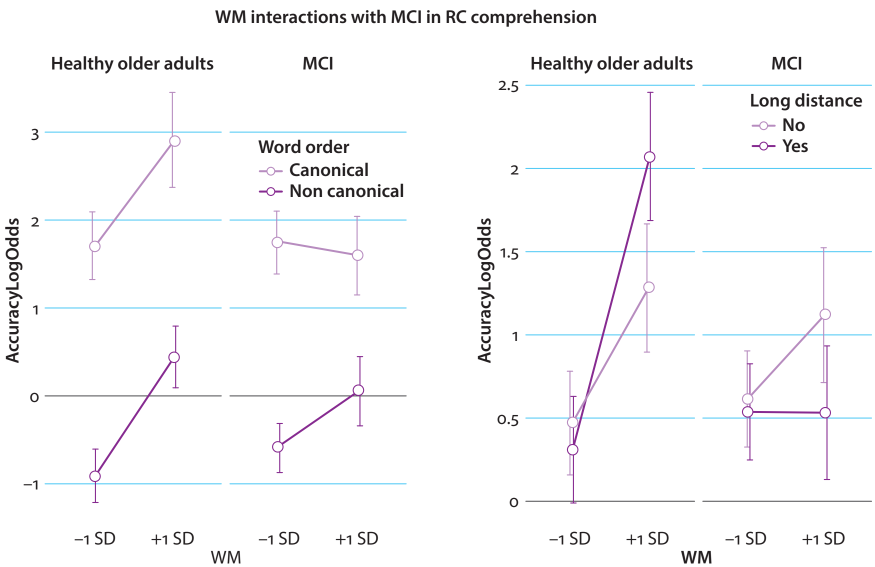

Figure 5. Interactions of WM and sentence type
Note: Effects are estimated at ±1 SD of WM scores. Error bars represent the standard error of the parameter estimates.

In relation to embedded clauses, object-extraction led to more failures in comprehension than subject-extraction. WM capacity played an overall positive role in accuracy across groups. There was also a statistically significant interaction between WM and MCI, which revealed a lower effect of WM aiding to comprehension in participants with this neurological condition. Moreover, an interaction was found between long-distance integration, WM and presence of MCI. This was due to the fact that WM capacity was found to be helpful in the comprehension of the most complex ORst sentences for healthy older adults, but there was no positive relationship between the WM capacity and the comprehension of these sentences in participants with MCI. These effects are displayed in Figure 5. With regard to the comprehension of the content related to the main clause, the only effect of interest was a slight lower accuracy in older adults compared to young adults.

Relationship among reading times and comprehension accuracy
As a complementary aim, we wanted to test relationships between on-line processing times and comprehension performance for our different groups. As stated above, longer processing times associated with comprehension hits would mean that participants use their cognitive skills to provide additional processing time to compensate for difficulties. On the contrary, longer processing times associated with comprehension failures would mean that longer processing times are associated with cognitive difficulties. However, it is also possible that those patterns might be blurred and masked by other factors and processes mediating between on-line reading and post-sentence task performance, so that no relationship between processing time and accuracy was detected.

Different generalized linear mixed-effect models were fitted for each group considering the possible interaction effects between residual reading times, sentence structure and cognitive skills over the success on comprehension probes. Given that comprehension difficulties depending on sentence structure were noticed only for thematic assignment in the RC and taking into account that the initial and final regions of the RC were regions where cognitive skills mainly interacted with sentence structure and group, we focus our analysis on these regions. Further, we only considered items carrying verification probes about thematic assignment of the RC.

Hence, each model considered the main and interaction effects on post-sentence verification probe accuracy of (a) the processing time at that region, as reflected by residual reading times, (b) sentence structure (using Helmert coding to contrast effects of non-canonical word order and long-distance integration, see Table 4), and (c) WM and IC (scaled to z scores). Different models were fitted for each group (young adults, healthy older adults and people with MCI). Because these analyses were mainly exploratory, the low number of items questioning

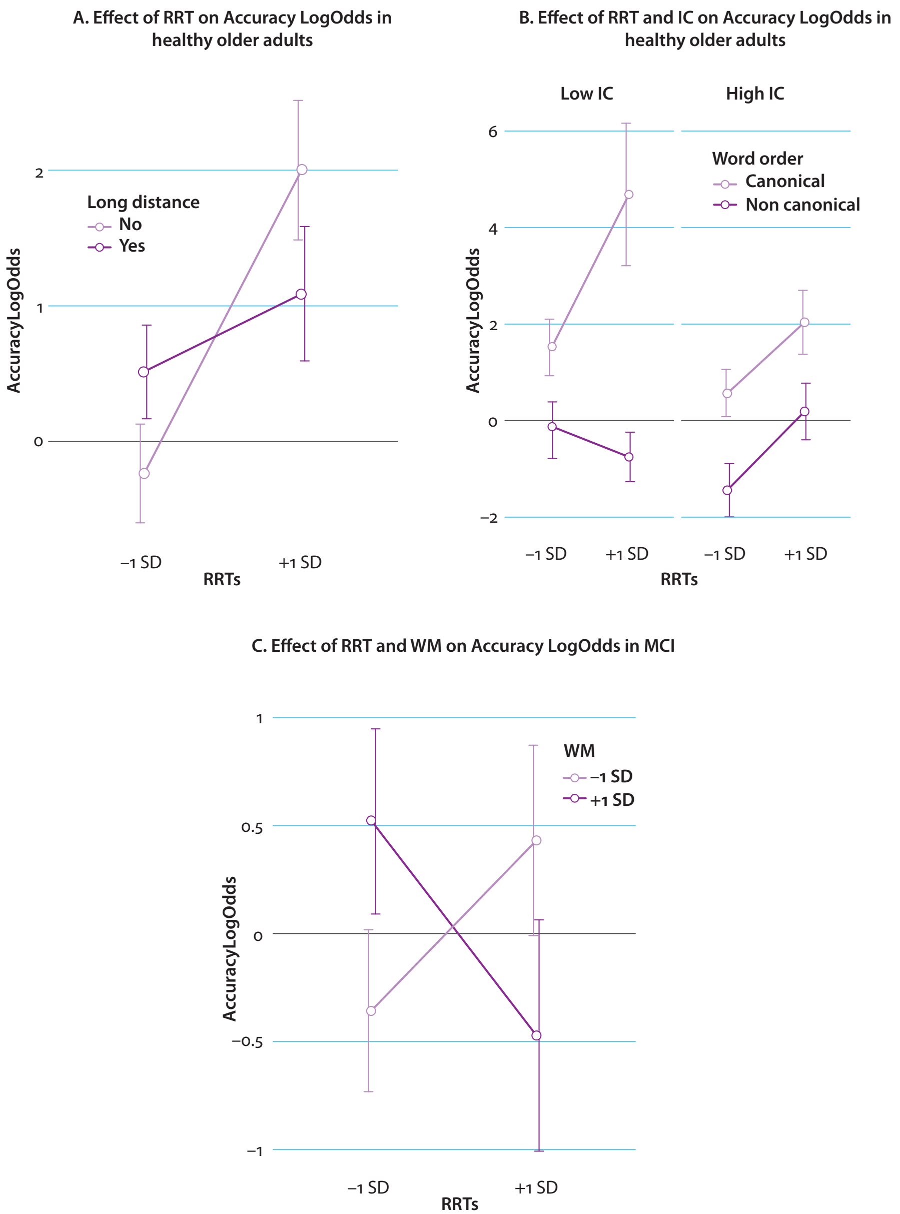

Figure 6. Effect of residual reading time (RRT) at RC closure on comprehension accuracy in older adults
Note: Parameters are estimated at ±1 SD of RRTs. Error bars represent the standard error of the parameter estimates.

about the content of RCs, and to ensure model convergence, only random intercepts by subjects and items were considered. To summarize the results, residual reading times were related to accuracy only in healthy older adults and people with MCI, and only at RC closure. These data are presented on Appendix 3. Figure 6 displays the effects of interest.

According to these analyses, processing time, as expressed by residual reading times, had little impact on determining the subsequent accuracy on end-of sentence verification probes in young adults. However, the results were different for the group of healthy older adults. In this group, residual reading times at RC closure were positively associated with off-line comprehension accuracy for embedded clauses. This effect was attenuated for the most complex ORst sentences, as indicated by the negative interaction of residual reading times and long-distance (see Figure 6a). Furthermore, the beneficial effect of residual reading times was different in healthy older adults with low IC skills. In this group, increased residual reading times were beneficial for the comprehension of canonical sentences, but enlarged residual reading times in healthy older adults with higher IC skills characterized non-canonical sentences which participants failed to comprehend (see Figure 6b). Finally, with regard to the participants with MCI, a negative association of residual reading times at RC closure and WM with off-line accuracy was found (see Figure 6c).

4. Discussion

The aim of the present research was to investigate influences of WM and IC in sentence processing across young adults and older adults with and without MCI. Interestingly, the present results found evidence suggesting a modulatory role of WM and IC during sentence reading for older adults, depending on the syntactic complexity of the material, and also on the presence of MCI.

In a nutshell, the results show that all groups were sensitive to sentence complexity while reading the RC. An increase in reading times was seen as soon as readers found the relative pronoun region. In the current study, non-canonical RCs were early marked as such by the Spanish preposition "a" (del Río et al., 2012). Subsequent syntactic complexity effects on RC reading times were modulated by WM capacity, age, and the presence of MCI. Effects of IC were also noted at the end of the RC and at the immediately adjacent main clause verb and depended mainly on ageing. Finally, accuracy on the off-line comprehension task was lower for non-canonical sentences and was strongly influenced by individual WM capacity across young adults and healthy older adults. However, the end-of-sentence comprehension performance of participants with MCI was not aided by their WM capacity, particularly in the most complex sentence type.

Next, we will discuss the implications of effects concerning on-line processing for theoretical proposals on the role of WM and IC in sentence processing among older adults. We will finish by discussing effects on comprehension accuracy.

4.1 Modulatory effects of WM across the RC

We divided the RC region which immediately follows the relative pronoun into two successive segments. The initial region of the RC comprised the verb and the following article in SR and ORsi sentences, but the clause-internal subject NP in ORst sentences. At this region, residual reading times were faster for young adults in ORst sentences, and no differences appeared between SR and ORsi. This might happen because SR and ORsi sentences contain the embedded verb, triggering similar demands in relation to thematic role assignment. Conversely, ORst sentences contain the embedded subject NP at this sentence position. Hence, it seems that complex sentences require little enhanced processing at this point for young adults. However, the pattern was clearly different in older adults. In this population, non-canonical complex sentences (ORsi and ORst) engaged longer residual reading times, particularly among older adults with larger WM capacity.

A priori, one might expect that greater WM capacity would play a facilitative role in the processing of syntactic complexity. If so, older adults with higher WM skills should have shown shorter reading times for the complex sentences. However, the pattern we found was in fact the opposite: slower reading times for the older adults with higher WM scores. This finding appears to be in agreement with proposals made by MacDonald, Just, & Carpenter (1992), Nicenboim et al. (2016) or Stine-Morrow et al. (2000). According to these views, higher WM capacity might serve to compensate for difficulties with complex linguistic material by enabling additional processing, at the cost of longer processing times.

Caplan et al. (2011) found slower reading times positively associated with better comprehension results, although only in some specific sentence types. Interestingly, they suggested that increased reading times were related to the use of WM for strategic encoding of the material, not for parsing operations *per se*. This line of reasoning is quite congruent with our current results. Older adults (and, particularly, those with high WM capacity) slowed their reading times in non-canonical complex sentences when entering in the RC. But this initial region of the RC does not involve the establishment of long-distance dependencies between the embedded verb and extracted object, which supposedly might trigger WM demands for integration during on-line processing (Gibson, 1998; Lewis et al., 2006). In fact, young adults with higher WM dedicated less processing time at this point to complex sentences. Hence, the increase in processing time at this point does not seem to be justified by parsing operations *per se*. A plausible explanation might be that, as the RC of ORsi and ORst sentences is identified as non-canonical at the pronoun region, older adults use strategic adaptations to attempt to achieve an adequate comprehension performance (Stine-Morrow & Payne, 2016).

Interesting effects also came into play at the following region (RC closure). At this point larger WM capacity was again associated with longer residual reading times. It is interesting to note that this second region of the RC implies its closure, and it has been associated with wrap-up processes (i.e., integration of information across the clause). The effect of increased residual reading times interacted with WM in older adults, so that residual reading times were faster in older adults with low WM capacity. Moreover, the recruitment of processing time in the complex non-canonical sentences was larger for healthy older adults with good WM skills. Older adults with MCI and with good WM capacity showed also longer residual reading times, but not especially for non-canonical sentences.

The up-regulation of clause and sentence wrap-up effects have been repeatedly noted in older adults (Stine-Morrow & Payne, 2016). In fact, wrap-up effects have been found to tend to be longer for older adults with high WM (Payne et al., 2012). Recent research by Payne & Stine-Morrow (2016) shows that post-sentence recall of sentence content is directly proportional to end-of-sentence wrap-up in healthy older adults. In contrast, older adults at risk for MCI show lower wrap-up effects and a lower recall of sentence information.

As already advanced, this line of argumentation predicts some degree of association between on-line reading times and comprehension performance in older adults. In particular, we would expect longer reading times to correlate positively with accuracy in complex sentences. The exploratory analysis concerning relationships among residual reading times and accuracy showed some positive evidence in this regard. Results from the end of the RC provided some evidence for a compensatory recruitment of processing time to aid comprehension in healthy older adults. Specifically, older adults with upregulating reading times at clause boundary seem to have a better level of comprehension. An additional interesting result is that the beneficial effect for comprehension of enlarged reading times was lower for sentences with long-distance dependencies (ORst, Figure 6a). In other words, there is a lower positive effect of longer reading times on comprehension accuracy for the most complex kind of sentences. This might be due either to the fact that the up-regulation of reading time is not as efficient in compensating for processing difficulties in these more complex sentences, or to other difficulties happening in the interval between the RC and the off-line probe masking this relationship. Anyway, even if the positive effect of reading times for sentence comprehension in healthy older adults was attenuated in sentences with long-distance dependencies, it did not disappear completely.

The modulatory effect of WM for non-canonical sentences was lower in participants with MCI. Therefore, this group seems to show a lower capacity to recruit additional reading times to facilitate comprehension for complex sentences. Moreover, the exploratory analysis relating processing time and accuracy showed that longer residual reading times were associated with failures in comprehension among those older adults with MCI who had better WM capacity. This suggests that the recruitment of additional reading time is not enough for adequate compensation. Akin to the current results, Payne & Stine-Morrow (2016) not only found that older adult at-risk for amnestic MCI had lower end-of-sentence wrap-up effects but also that, in contrast with the effect in healthy older adults, enlarged wrap-up was not associated with better memory for sentence information.

In summary, our results are congruent, at least to some extent, with the idea that older adults recruit WM resources in a compensatory manner, and this happened during the processing of the embedded RC. However, the reading times of complex sentences in older adults with MCI are less influenced by WM capacity and enlarged processing times were not clearly associated with a better sentence comprehension in this population.

4.2 Modulatory effects of IC at RC closure and main verb

Cognitive control of conflicting and/or interfering information is negatively influenced by older age (Hasher et al., 2007), and it is disrupted to a higher extent in people with MCI (Aurtenetxe et al., 2016; Traykov et al., 2007). In the present study, participants might have experienced a conflicting situation while processing non-canonical RCs. This happens because the initial, animate NP of sentences prompts a sentence interpretation with such NP as subject and agent. This thematic role assignment is valid for the main clause (and should be retained until arriving at the main clause verb). However, when facing a non-canonical object-extracted RC, this interpretation should be inhibited, and such inhibition might require the recruitment of cognitive control resources. The interaction of IC, sentence complexity and age suggest that the interference effects associated with non-canonical sentences influence older adults to a higher extent than young participants.

The current sample of healthy older adults did not show significant differences in IC, as measured by the Stroop index, when compared to the group of young participants. This fact might call into question the assumption of age-related IC difficulties in this group. Notwithstanding, previous neuroimaging studies have found that older adults might show a similar behavioral performance than young adults in the Stroop task (Langenecker, Nielson, & Rao, 2004), but this similar behavioral performance of young and older adults is maintained at the expense of an increased recruitment of inhibitory-related areas of the frontal lobe. In turn, older adults with MCI participating in the present study showed a clear decline in the IC Stroop index when compared to young adults and healthy older adults. This result evidences the progression of frontal deficits in the current sample of older adults with MCI.

IC skills have been shown to play an important role in language comprehension when it is necessary to manage conflicting interpretations (Hsu & Novick, 2016; Vuong & Martin, 2014), and this might be a good example of such situations. Studies in older adults have also shown how comprehension is greatly influenced by their IC skills (Goral et al., 2011; Yoon et al., 2015). The current results endorse those of previous studies, and additionally find that the complexity in terms of conflict posed by non-canonical word order might be detected on-line when arriving at the closure of the RC. Moreover, this effect of IC on reading times seems to hold for both groups of older adults, as it was associated with the contrast based on age in the linear mixed-effect model, but there were no significant effects regarding the contrast based on MCI.

The analysis exploring possible relationships between residual reading times and end-of-sentence comprehension showed that the beneficial wrap-up effect for RC comprehension seen in healthy older adults is modulated by IC skills. In healthy older adults with high IC skills, slower residual reading times were associated with a better comprehension performance, in congruence with the beneficial effect of clause wrap-up. This beneficial effect for comprehension occurs also in healthy older adults with low IC for the simple, canonical sentences. Crucially, however, enlarged residual reading times in non-canonical sentences are associated with the likelihood of comprehension failures in older people with low IC (Figure 6b). This result seems congruent with the "longer processing time means disrupted processing" view, but this applies particularly to the role of IC on non-canonical sentences. When confronted with the non-canonical sentence structures, older adults with low IC skills might be unable to suppress the canonical interpretation in several trials. This leads to increased conflict, rising processing times and resulting in a failed suppression of the incorrect thematic assignment. These results emphasize how the possible loss of IC skills with age might contribute to burden language processing in older adults.

The same effect is not detected when exploring the relationships between residual reading times and end-of-sentence comprehension in MCI. There might be several reasons for this, but the fact that older adults with MCI are very affected by WM difficulties might have obscured any relationship between IC, RRT and comprehension accuracy. The possible interplay of effects of WM and IC on sentence comprehension for people with MCI warrants further research.

The interaction effect between IC and sentence complexity rising at the end of the RC in older adults is inverted at the following main clause verb. Here, good IC skills lead to longer residual reading times at the non-canonical sentences in older adults. This result might seem counterintuitive at a first glance. However, it is not so surprising if we consider that, for non-canonical sentences, the reader has just inhibited, supposedly, the interpretation of the first NP as subject and agent for the RC. But when turning to the main clause again, the first NP is in fact the subject and agent of the sentence (i.e., there is a perspective shifting, see MacWhinney & Pléh, 1988). Hence, akin to a negative priming effect (Tipper, 2001), once those older adults with better IC skills solve interference by the initial NP as subject of the RC, they might also need more time to retrieve back the initial NP as subject of the main clause when arriving at the main verb. However, this is just a tentative explanation and this rationale should be explored in a greater detail by future studies.

In summary, it seems that both WM and IC contribute to sentence processing in older adults but with different roles. On one side, higher WM capacity seems to provide resources to strategically adapt reading times, with more or less success, while IC is used to manage conflicting interpretations arising during language processing.

4.3 Effects on sentence comprehension

The study of end-of-sentence comprehension also shows some interesting results. Difficulties in sentence comprehension were mainly noticed for non-canonical embedded RCs, confirming that participants were challenged by thematic assignment when the object preceded the subject. The main determinant of the level of sentence comprehension in the embedded RCs was the individual WM capacity. In fact, while no differences in comprehension accuracy could be mainly attributed to age and presence of MCI, a main effect and several interactions were detected with regard to WM. Hence, given that all three groups significantly differed in WM, it might be assumed that such differences accounted for an important amount of the variance in comprehension accuracy across groups.

Contrary to the clear influence of IC found on processing time among older adults, there was no clear effect of IC in comprehension accuracy. This result contrasts with those from previous studies which analyzed off-line the effect of cognitive control in sentence comprehension (Goral et al., 2011; Yoon et al., 2015). However, it would be premature to deem IC as irrelevant in sentence comprehension according to the current results. As discussed in the previous section, exploratory analysis on the relationship between residual reading times and accuracy show that IC modulated the relationship of on-line reading times with post-sentence comprehension in healthy older adults.

Notwithstanding, it seems obvious that WM plays a prominent role in off-line sentence comprehension across all groups. If we consider how the task develops, the strong role played by WM on accuracy of post-sentence verification probes is not unexpected. Because participants first proceed to read the sentence, and then are shown the verification probe, off-line performance is heavily dependent on the capacity to form good quality memory traces and retrieve them in an efficient manner (Unsworth & Engle, 2007).

It is interesting to note how participants with MCI have a lower positive effect of WM. This parallels the lack of WM modulation which is seen during on-line processing in participants with MCI as compared to healthy older adults. Moreover, in this group, the lack of a positive effect of WM for comprehension is particularly noticeable for the most complex ORst clauses. As older adults with MCI are more challenged by WM difficulties than healthy older adults, the pattern of results shown by participants with MCI suggest that their limitations in WM capacity did not allow them to benefit from WM to assist the comprehension task. In other words, it is likely that participants with MCI are overwhelmed in their capacity to use WM to aid comprehension, this fact being even more obvious for the most complex ORst sentences.

5. Conclusions

The current results have several limitations. One example is the low number of participants, particularly if the group of participants with MCI is considered. Furthermore, there was substantial heterogeneity in this group, with both single- and multi-domain amnestic MCI diagnosis. Another problem concerning the current sample is the difference in years of schooling between young and older adults, which we have tried to address by using years of schooling as a covariate in our analysis. Older adults with MCI have also more depressive complaints than healthy older adults, although we excluded participants with depressive symptomatology. In any case, further studies should investigate in greater detail the implications for sentence processing and comprehension of the wide spectrum of age-related cognitive changes.

Self-paced reading is widely accepted as capturing, with enough sensitivity, the on-line processing of language. But it is also well established that more detailed information can be gathered by using eye-tracking while reading or the visual world paradigm (e.g., Kemper & Liu, 2007; Sekerina, Campanelli, & Van Dyke, 2016). Therefore, it would be interesting to try to replicate the current results using other complementary techniques sensitive to the on-line processes involved in sentence comprehension.

Finally, measures of WM and IC were established by using a single task for each domain. While the OST and the Golden version of the Stroop task are widely considered good measures of WM and IC, respectively, some researchers propose the use of composite scores from several tasks which seek to measure these constructs (e.g., Waters & Caplan, 2002). It might be a good idea to incorporate more comprehensive measures of WM and IC to future studies on this topic. In particular, IC is a complex construct and several inhibition-related functions have been proposed, such as resistance to proactive memory interference, inhibition of proponent responses or control of distractor interference (Friedman & Miyake, 2004). Moreover, their differential susceptibility to age-related changes are under active research (Pettigrew & Martin, 2014; Rey-Mermet & Gade, 2018), so that future studies might consider in a wide detail the consequences of age-related changes in different components of IC for language processing.

Notwithstanding these limitations, the current study provides suggestive evidence about how individual differences in WM and IC across older adults influence the on-line processing of syntactically complex sentences. Variability in IC skills across older adults seems to influence how they cope with conflict when different interpretations arise on-line during the course of sentence processing. Moreover, healthy older adults might strategically recruit additional processing time depending on their WM capacity for the processing of syntactically complex sentences, although with limited success for the most complex structures, while older adults with MCI were particularly challenged by WM limitations during sentence processing and comprehension.

Acknowledgements

The authors would like to acknowledge the help we received from Raquel Fernández Sánchez and Marta Galindo Fuentes on data collection and the collaboration of personnel from the Center for Cognitive Impairment Prevention at Madrid. We also want to thank to Professor John Grinstead from The Ohio State University (Columbus, USA) for his help with the English version of this manuscript.

Funding

This study was supported by the Spanish Ministry of Economy of Competitiveness under Grants PSI2012-38375-C03-03 and PSI2015-68793-C3-3.

References

Artero, S., Petersen, R., Touchon, J., & Ritchie, K. (2006). Revised criteria for mild cognitive impairment: Validation within a longitudinal population study. *Dementia and Geriatric Cognitive Disorders*, 22(5–6), 465–470. https://doi.org/10.1159/000096287

Aurtenetxe, S., García-Pacios, J., del Río, D., López, M. E., Pineda-Pardo, J. A., Marcos, A., … Maestú, F. (2016). Interference impacts working memory in mild cognitive impairment. *Frontiers in Neuroscience*, 10(OCT). https://doi.org/10.3389/fnins.2016.00443

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 63(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Baudiffier, V., Caplan, D., Gaonac'h, D., & Chesnet, D. (2011). The effect of noun animacy on the processing of unambiguous sentences: Evidence from French relative clauses. *The Quarterly Journal of Experimental Psychology*, 64(10), 1896–1905. https://doi.org/10.1080/17470218.2011.608851

Betancort, M., Carreiras, M., & Sturt, P. (2009). The processing of subject and object relative clauses in Spanish: An eye-tracking study. *The Quarterly Journal of Experimental Psychology*, 62(10), 1915–1929. https://doi.org/10.1080/17470210902866672

Bickel, C., Pantel, J., Eysenbach, K., & Schröder, J. (2000). Syntactic comprehension deficits in Alzheimer's disease. *Brain and Language*, 71(3), 432–448. https://doi.org/10.1006/brln.1999.2277

Borella, E., Carretti, B., & De Beni, R. (2008). Working memory and inhibition across the adult life-span. *Acta Psychologica*, 128(1), 33–44. https://doi.org/10.1016/j.actpsy.2007.09.008

Caplan, D., DeDe, G., Waters, G., Michaud, J., & Tripodis, Y. (2011). Effects of age, speed of processing, and working memory on comprehension of sentences with relative clauses. *Psychology and Aging*, 26(2), 439–450. https://doi.org/10.1037/a0021837

Caplan, D., & Waters, G. S. (2013). Memory mechanisms supporting syntactic comprehension. *Psychonomic Bulletin & Review*, 20(2), 243–268. https://doi.org/10.3758/s13423-012-0369-9

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioural and Brain Sciences*, 22, 77–126. https://doi.org/10.1017/S0140525X99001788

Casado, P., Martín-Loeches, M., Muñoz, F., & Fernandez-Frías, C. (2005). Are semantic and syntactic cues inducing the same processes in the identification of word order? *Cognitive Brain Research*, 24(3), 526–543. https://doi.org/10.1016/j.cogbrainres.2005.03.007

Chafetz, M. D., & Matthews, L. H. (2004). A new interference score for the Stroop test. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 19(4), 555–567. https://doi.org/10.1016/j.acn.2003.08.004

Croot, K., Hodges, J. R., & Patterson, K. (1999). Evidence for impaired sentence comprehension in early Alzheimer's disease. *Journal of the International Neuropsychological Society*, 5(5), 393–404. https://doi.org/10.1017/S1355617799555021

De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-3184-1

De Vincenzi, M. (1996). Syntactic analysis in sentence comprehension: Effects of dependency types and grammatical constraints. *Journal of Psycholinguistic Research*, 25(1), 117–133. https://doi.org/10.1007/BF01708422

DeCaro, R., Peelle, J. E., Grossman, M., & Wingfield, A. (2016). The two sides of sensory-cognitiveiInteractions: Effects of age, hearing acuity, and working memory span on sentence comprehension. *Frontiers in Psychology*, 7(February), 236. https://doi.org/10.3389/fpsyg.2016.00236

DeDe, G., Caplan, D., Kemtes, K., & Waters, G. (2004). The relationship between age, verbal working memory, and language comprehension. *Psychology and Aging*, 19(4), 601–616. https://doi.org/10.1037/0882-7974.19.4.601

del Río, D., López-Higes, R., & Martín-Aragoneses, M. T. M. T. (2012). Canonical word order and interference-based integration costs during sentence comprehension: The case of Spanish subject- and object-relative clauses. *Quarterly Journal of Experimental Psychology (2006)*, 65(11), 2108–2128. https://doi.org/10.1080/17470218.2012.674951

del Río, D., Maestú, F., López-Higes, R., Moratti, S., Gutiérrez, R., Maestú, C., & del-Pozo, F. (2011). Conflict and cognitive control during sentence comprehension: Recruitment of a frontal network during the processing of Spanish object-first sentences. *Neuropsychologia*, 49(3), 382–391. https://doi.org/10.1016/j.neuropsychologia.2010.12.005

Di Domenico, A., & Di Matteo, R. (2009). Processing Italian relative clauses: Working memory span and word order effects on RTs. *The Journal of General Psychology*, 136(4), 387–406. https://doi.org/10.1080/00221300903266671

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Golden, C. J. (2001). *Test de colores y palabras Stroop manual*. Madrid: TEA.

Goral, M., Clark-Cotton, M., Spiro, A., Obler, L. K., Verkuilen, J., & Albert, M. L. (2011). The contribution of set Switching and working memory to sentence processing in older adults. *Experimental Aging Research*, 37(5), 516–538. https://doi.org/10.1080/0361073X.2011.619858

Grodner, D., & Gibson, E. (2005). Some consequences of the serial nature of linguistic input. *Cognitive Science*, 29(2), 261–290. https://doi.org/10.1207/s15516709cog0000_7

Grossman, M., & Rhee, J. (2001). Cognitive resources during sentence processing in Alzheimer's disease. *Neuropsychologia*, 39(13), 1419–1431. https://doi.org/10.1016/S0028-3932(01)00059-8

Hartman, M., & Hasher, L. (1991). Aging and suppression: Memory for previously relevant information. *Psychology and Aging*, 6(4), 587–594. https://doi.org/10.1037/0882-7974.6.4.587

Hasher, L., Lustig, C., & Zacks, R. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195168648.003.0009

Homack, S., & Riccio, C. A. (2004). A meta-analysis of the sensitivity and specificity of the Stroop Color and Word Test with children. *Archives of Clinical Neuropsychology : The Official Journal of the National Academy of Neuropsychologists*, 19(6), 725–743. https://doi.org/10.1016/j.acn.2003.09.003

Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates Recovery from misinterpretation during real-time language processing. *Psychological Science*, 27(4), 572–582. https://doi.org/10.1177/0956797615625223

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., Thulborn, K. R., & Just, M. A. (1996). Brain activation modulated by sentence comprehension. *Science*, 274, 114–116. https://doi.org/10.1126/science.274.5284.114

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology. General*, 111(2), 228–238. https://doi.org/10.1037/0096-3445.111.2.228

Kemper, S., & Liu, C.-J. (2007). Eye movements of young and older adults during reading. *Psychology and Aging*, 22(1), 84–93. https://doi.org/10.1037/0882-7974.22.1.84

Kutas, M., & King, J. W. (1995). Who did what and when? Using word-and clause-Level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7(3), 376–395. https://doi.org/10.1162/jocn.1995.7.3.376

Langenecker, S. A., Nielson, K. A., & Rao, S. M. (2004). fMRI of healthy older adults during Stroop interference. *NeuroImage*, 21(1), 192–200. https://doi.org/10.1016/j.neuroimage.2003.08.027

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454. https://doi.org/10.1016/j.tics.2006.08.007

Lobo, A., Saz, P., Marcos, G., Día, J. L., de la Cámara, C., Ventura, T., Aznar, S. (1999). Revalidation and standardization of the cognition mini-exam (first Spanish version of the Mini-Mental Status Examination) in the general geriatric population. *Medicina Clinica*, 112(20), 767–774.

López-Higes, R., Prados, J. M., Rubio, S., Montejo, P., & Del Río, D. (2016). Executive functions and linguistic performance in SCD older adults and healthy controls. *Aging, Neuropsychology, and Cognition*, 14(00), 1–18. https://doi.org/10.1080/13825585.2016.1256370

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24(1), 56–98. https://doi.org/10.1016/0010-0285(92)90003-K

MacWhinney, B., & Pléh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, 29(2), 95–141. https://doi.org/10.1016/0010-0277(88)90034-0

Mak, W. M., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language*, 54(4), 466–490. https://doi.org/10.1016/j.jml.2006.01.001

Marková, J., Horváthová, Ľ., Králová, M., & Cséfalvay, Z. (2017). Sentence comprehension in Slovak-speaking patients with Alzheimer's disease. *International Journal of Language & Communication Disorders*, 52(4), 456–468. https://doi.org/10.1111/1460-6984.12284

Nicenboim, B., Logačev, P., Gattei, C., & Vasishth, S. (2016). When high-capacity readers slow down and low-capacity readers speed up: Working memory and locality effects. *Frontiers in Psychology*, 7(MAR), 280. https://doi.org/10.3389/fpsyg.2016.00280

Payne, B. R., Gao, X., Noh, S. R., Anderson, C. J., & Stine-Morrow, E. A. L. (2012). The effects of print exposure on sentence processing and memory in older adults: Evidence for efficiency and reserve. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, 19(1–2), 122–149. https://doi.org/10.1080/13825585.2011.628376

Payne, B. R., & Stine-Morrow, E. A. L. (2016). Risk for mild cognitive impairment is associated with semantic integration deficits in sentence processing and memory. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, 71(2), 243–253. https://doi.org/10.1093/geronb/gbu103

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256(3), 183–194. https://doi.org/10.1111/j.1365-2796.2004.01388.x

Pettigrew, C., & Martin, R. C. (2014). Cognitive declines in healthy aging: Evidence from multiple aspects of interference resolution. *Psychology and Aging*, 29(2), 187–204. https://doi.org/10.1037/a0036085

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reisberg, B., Ferris, S. H., De Leon, M. J., & Crook, T. (1982). The global deterioration scale for assessment of primary degenerative dementia. *American Journal of Psychiatry*, 139(9), 1136–1139. https://doi.org/10.1176/ajp.139.9.1136

Rey-Mermet, A., & Gade, M. (2018). Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin and Review*, 25(5), 1695–1716. https://doi.org/10.3758/s13423-017-1384-7

Rogalsky, C., Matchin, W., & Hickok, G. (2008). Broca's area, sentence comprehension, and working memory: An fMRI Study. *Frontiers in Human Neuroscience*, 2(October), 14. https://doi.org/10.3389/neuro.09.014.2008

See, S. T., & Ryan, E. B. (1995). Cognitive mediation of adult age differences in language performance. *Psychology and Aging*, 10(3), 458–468. https://doi.org/10.1037/0882-7974.10.3.458

Sekerina, I. A., Campanelli, L., & Van Dyke, J. A. (2016). Using the visual world paradigm to study retrieval interference in spoken language comprehension. *Frontiers in Psychology*, 7(JUN), 1–15. https://doi.org/10.3389/fpsyg.2016.00873

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS) recent evidence and eevelopment of a shorter version. *Clinical Gerontologist*, 5(1–2), 165–173. https://doi.org/10.1300/J018v05n01_09

Small, J. A., Kemper, S., & Lyons, K. (2000). Sentence repetition and processing resources in Alzheimer's disease. *Brain and Language*, 75(2), 232–258. https://doi.org/10.1006/brln.2000.2355

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, 22(2), 461–479. https://doi.org/10.1037/0096-1523.22.2.461

Stine-Morrow, E. A. L., & Payne, B. R. (2016). Age differences in language segmentation. *Experimental Aging Research*, 42(1), 83–96. https://doi.org/10.1080/0361073X.2016.1108751

Stine-Morrow, E. A. L., Ryan, S., & Leonard, J. S. (2000). Age differences in on-line syntactic processing. *Experimental Aging Research*, 26(December), 315–322. https://doi.org/10.1080/036107300750015714

Taler, V., & Phillips, N. A. (2008). Language performance in Alzheimer's disease and mild cognitive impairment: A comparative review. *Journal of Clinical and Experimental Neuropsychology*, 30(5), 501–556. https://doi.org/10.1080/13803390701550128

Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 54(2), 321–343. https://doi.org/10.1080/713755969

Traykov, L., Raoux, N., Latour, F., Gallo, L., Hanon, O., Baudic, S., … Rigaud, A.-S. (2007). Executive functions deficit in mild cognitive impairment. *Cognitive and Behavioral Neurology*, 20(4), 219–224. https://doi.org/10.1097/WNN.0b013e31815e6254

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*. https://doi.org/10.1006/jmla.1994.1014

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114(1), 104–132. https://doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498–505. https://doi.org/10.3758/BF03192720

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 33(2), 407–430. https://doi.org/10.1037/0278-7393.33.2.407

Vuong, L. C., & Martin, R. C. (2014). Domain-specific executive control and the revision of misinterpretations in sentence comprehension. *Language, Cognition and Neuroscience*, 29(3), 312–325. https://doi.org/10.1080/01690965.2013.836231

Waters, G., & Caplan, D. (2002). Working memory and online syntactic processing in Alzheimer's disease. *Journal of Gerontology: Psychological Sciences*, 57B(4), 298–311. https://doi.org/10.1093/geronb/57.4.P298

Wechsler, D. (1997). *Wechsler Memory Scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.-O., … Petersen, R. C. (2004). Mild cognitive impairment–beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*, 256(3), 240–246. https://doi.org/10.1111/j.1365-2796.2004.01380.x

Yoon, J., Campanelli, L., Goral, M., Marton, K., Eichorn, N., & Obler, L. K. (2015). The effect of plausibility on sentence comprehension among older adults and its relation to cognitive functions. *Experimental Aging Research*, 41(3), 272–302. https://doi.org/10.1080/0361073X.2015.1021646

Appendix 1. Results of linear mixed effect models for on-line reading times

Supplementary Table 1.1 Summary of fixed effects parameters at RC pronoun

	Estimate	**S.E.**	***t*-value**
(Intercept)	−0.0128	0.0109	−1.18
Non-canonical	**0.0833**	**0.0212**	**3.93**
AGE	**−0.1090**	**0.0224**	**−4.88**
MCI	−0.0464	0.0255	−1.82
WM	−0.0064	0.0074	−0.86
IC	0.0019	0.0065	0.29
Schooling	**0.0183**	**0.0069**	**2.64**
Non-canonical:AGE	0.0538	0.0512	1.05
Non-canonical:MCI	0.0688	0.0584	1.18
Non-canonical:WM	0.0015	0.0170	0.09
Non-canonical:IC	0.0021	0.0149	0.14
Non-canonical:Schooling	−0.0089	0.0158	−0.56
AGE:WM	−0.0212	0.0182	−1.16
AGE:IC	0.0058	0.0161	0.36
AGE:Schooling	0.0171	0.0211	0.81
MCI:WM	−0.0173	0.0203	−0.85
MCI:IC	0.0035	0.0176	0.20
MCI:Schooling	0.0087	0.0154	0.56
Non-canonical:AGE:WM	0.0437	0.0418	1.04
Non-canonical:AGE:IC	0.0237	0.0369	0.64
Non-canonical:AGE:Schooling	0.0342	0.0482	0.71
Non-canonical:MCI:WM	−0.0045	0.0466	−0.10
Non-canonical:MCI:IC	0.0019	0.0404	0.05
Non-canonical:MCI:Schooling	−0.0116	0.0353	−0.33

Supplementary Table 1.2 Summary of fixed effects parameters across RC and main verb

	Initial region of RC			RC closure			Main verb		
	Estimate	S.E.	*t*-value	Estimate	S.E.	*t*-value	Estimate	S.E.	*t*-value
(Intercept)	−0.0028	0.0134	−0.21	**0.0676**	**0.0270**	**2.50**	**0.0856**	**0.0184**	**4.64**
Non-canonical	0.0369	0.0232	1.59	0.0473	0.0360	1.31	0.0347	0.0293	1.18
Long-distance	−0.0197	0.0182	−1.09	−0.0280	0.0278	−1.01	0.0404	0.0253	1.60
AGE	−0.0377	0.0293	−1.29	−0.0706	0.0618	−1.14	**−0.1126**	**0.0394**	**−2.86**
MCI	−0.0025	0.0334	−0.08	−0.0200	0.0704	−0.28	**−0.0961**	**0.0448**	**−2.15**
WM	0.0123	0.0097	1.27	**0.0599**	**0.0205**	**2.93**	−0.0086	0.0130	−0.66
IC	−0.0020	0.0085	−0.24	−0.0129	0.0179	−0.72	0.0191	0.0114	1.68
Schooling	−0.0001	0.0090	−0.01	−0.0127	0.0190	−0.67	0.0123	0.0121	1.02
Non-canonical:AGE	**0.1451**	**0.0560**	**2.59**	0.0590	0.0872	0.68	−0.0701	0.0713	−0.98
Non-canonical:MCI	0.0123	0.0640	0.19	−0.1960	0.0991	−1.98	0.0652	0.0806	0.81
Long-distance:AGE	0.0028	0.0439	0.06	−0.0063	0.0672	−0.09	−0.0529	0.0614	−0.86
Long-distance:MCI	−0.0359	0.0500	−0.72	−0.1087	0.0763	−1.42	0.0001	0.0694	0.00
Non-canonical:WM	0.0367	0.0185	1.98	0.0242	0.0288	0.84	−0.0349	0.0235	−1.49
Non-canonical:IC	−0.0200	0.0162	−1.23	−0.0260	0.0251	−1.04	**0.0461**	**0.0206**	**2.24**
Non-canonical:Schooling	0.0172	0.0173	0.99	0.0206	0.0267	0.77	−0.0327	0.0219	−1.49
Long-distance:WM	**0.0322**	**0.0146**	**2.21**	−0.0054	0.0225	−0.24	0.0248	0.0202	1.23
Long-distance:IC	−0.0164	0.0127	−1.29	−0.0070	0.0195	−0.36	−0.0158	0.0176	−0.89
Long-distance:Schooling	**−0.0285**	**0.0135**	**−2.11**	−0.0113	0.0206	−0.55	**−0.0460**	**0.0189**	**−2.43**
AGE:WM	0.0260	0.0238	1.09	**0.1086**	**0.0504**	**2.16**	0.0025	0.0320	0.08
AGE:IC	0.0056	0.0210	0.27	−0.0623	0.0443	−1.41	0.0096	0.0282	0.34
AGE:Schooling	0.0086	0.0276	0.31	0.0020	0.0583	0.03	0.0377	0.0373	1.01
MCI:WM	−0.0146	0.0265	−0.55	−0.0257	0.0559	−0.46	−0.0260	0.0355	−0.73
MCI:IC	0.0101	0.0230	0.44	0.0119	0.0485	0.25	−0.0341	0.0308	−1.11
MCI:Schooling	−0.0015	0.0202	−0.07	−0.0163	0.0424	−0.38	−0.0250	0.0270	−0.93
Non-canonical:AGE:WM	**0.0917**	**0.0456**	**2.01**	−0.0165	0.0714	−0.23	−0.0298	0.0581	−0.51
Non-canonical:AGE:IC	−0.0467	0.0402	−1.16	**−0.1916**	**0.0623**	**−3.08**	**0.1534**	**0.0513**	**2.99**
Non-canonical:AGE:Schooling	0.0177	0.0529	0.33	0.0466	0.0821	0.57	−0.0186	0.0676	−0.28
Non-canonical:MCI:WM	−0.0434	0.0509	−0.85	**−0.1753**	**0.0789**	**−2.22**	0.0054	0.0643	0.08
Non-canonical:MCI:IC	−0.0040	0.0440	−0.09	0.0372	0.0676	0.55	−0.0024	0.0556	−0.04
Non-canonical:MCI:Schooling	−0.0022	0.0386	−0.06	−0.0311	0.0591	−0.53	−0.0010	0.0485	−0.02
Long-distance:AGE:WM	**0.0967**	**0.0359**	**2.70**	−0.0456	0.0556	−0.82	0.0221	0.0500	0.44
Long-distance:AGE:IC	−0.0402	0.0315	−1.28	−0.0463	0.0485	−0.96	−0.0525	0.0439	−1.20

Supplementary Table 1.2 (*continued*)

	Initial region of RC			RC closure			Main verb		
	Estimate	S.E.	*t*-value	Estimate	S.E.	*t*-value	Estimate	S.E.	*t*-value
Long-distance:AGE: Schooling	0.0010	0.0413	0.02	−0.0554	0.0633	−0.87	0.0751	0.0584	1.29
Long-distance:MCI:WM	−0.0293	0.0400	−0.73	−0.0465	0.0616	−0.75	−0.0771	0.0551	−1.40
Long-distance:MCI:IC	0.0346	0.0345	1.01	0.0169	0.0526	0.32	−0.0040	0.0475	−0.08
Long-distance:MCI: Schooling	−0.0129	0.0302	−0.43	0.0134	0.0457	0.29	0.0518	0.0416	1.24

Appendix 2. Results of generalized linear mixed effect models for off-line comprehension probes

Supplementary Table 2.1 Summary of fixed effects parameters regarding accuracy on verification probes

	Embedded RC			Main Clause		
	Estimate	S.E.	*z*-value	Estimate	S.E.	*z*-value
(Intercept)	**0.7670**	**0.2088**	**3.67*****	**1.7005**	**0.3274**	**5.19*****
Non-canonical	**−3.2042**	**0.4660**	**−6.87*****	−0.1160	0.4189	−0.27
Long-distance	−0.1398	0.3063	−0.45	−0.3030	0.3378	−0.89
AGE	−0.3149	0.3942	−0.79	**−0.8970**	**0.4573**	**−1.96***
MCI	−0.5017	0.4636	−1.08	−0.4900	0.5097	−0.96
WM	**0.4087**	**0.1267**	**3.22****	0.1049	0.1491	0.70
IC	0.0214	0.1090	0.19	0.1522	0.1312	1.16
Schooling	−0.0967	0.1148	−0.84	−0.1086	0.1402	−0.77
Non-canonical:AGE	−0.5349	1.0831	−0.49	0.6526	1.0240	0.63
Non-canonical:MCI	1.2954	1.2798	1.01	1.6733	1.1449	1.46
Long-distance:AGE	−0.4055	0.7199	−0.56	−1.3979	0.8362	−1.67
Long-distance:MCI	−1.2151	0.8587	−1.41	0.0101	0.9100	0.01
Non-canonical:WM	0.1522	0.3355	0.45	−0.1857	0.3406	−0.54
Non-canonical:IC	−0.0802	0.2914	−0.27	−0.3383	0.3008	−1.12
Non-canonical:Schooling	−0.3329	0.3032	−1.09	0.3172	0.3130	1.01
Long-distance:WM	0.1434	0.2391	0.59	−0.0264	0.2645	−0.09
Long-distance:IC	−0.2396	0.2047	−1.17	0.0676	0.2341	0.28
Long-distance:Schooling	−0.0289	0.2145	−0.13	−0.3474	0.2549	−1.36
AGE:WM	0.3081	0.3109	0.99	−0.2458	0.3715	−0.66

(*continued*)

Supplementary Table 2.1 (*continued*)

	Embedded RC			Main Clause		
	Estimate	**S.E.**	***z*-value**	**Estimate**	**S.E.**	***z*-value**
AGE:IC	−0.3041	0.2728	−1.11	−0.0896	0.3314	−0.27
AGE:Schooling	0.5254	0.3412	1.54	0.4136	0.4309	0.95
MCI:WM	**−0.7992**	**0.3455**	**−2.31***	0.1048	0.4043	0.25
MCI:IC	0.5740	0.2949	1.94	−0.4499	0.3521	−1.27
MCI:Schooling	0.0763	0.2666	0.28	0.1770	0.3109	0.56
Non-canonical:AGE:WM	−1.1710	0.8432	−1.38	1.1962	0.8518	1.40
Non-canonical:AGE:IC	1.0992	0.7422	1.48	−0.5121	0.7622	−0.67
Non-canonical:AGE:Schooling	−1.3762	0.8742	−1.57	0.1869	0.9491	0.19
Non-canonical:MCI:WM	0.6805	0.9105	0.74	0.8182	0.9299	0.87
Non-canonical:MCI:IC	−1.1001	0.7775	−1.41	0.6702	0.7988	0.83
Non-canonical:MCI:Schooling	0.1937	0.7290	0.26	−0.0594	0.7040	−0.08
Long-distance:AGE:WM	0.3887	0.5788	0.67	−0.1358	0.6726	−0.20
Long-distance:AGE:IC	−0.4413	0.5013	−0.88	0.7606	0.6028	1.26
Long-distance:AGE:Schooling	−0.2601	0.6456	−0.40	1.3388	0.7933	1.68
Long-distance:MCI:WM	**−1.3903**	**0.6632**	**−2.09***	−0.2670	0.7118	−0.37
Long-distance:MCI:IC	−0.0677	0.5551	−0.12	−0.2718	0.6188	−0.43
Long-distance:MCI:Schooling	−0.2435	0.4897	−0.49	0.1686	0.5535	0.30

***$p < .001$;
**$p < .01$;
*$p < .05$

Appendix 3. Results of generalized linear mixed effect models considering effects of reading times on comprehension accuracy

Supplementary Table 3.1 Summary of fixed effects parameters regarding effects of residual reading times (RRTs) over accuracy

	Healthy older adults			**MCI**		
	RC closure			**RC closure**		
	Estimate	**S.E.**	***z*-value**	**Estimate**	**S.E.**	***z*-value**
(Intercept)	**1.0580**	**0.3195**	**3.31*****	0.1895	0.4074	0.47
RRTs	**1.7875**	**0.7682**	**2.33***	−1.9549	1.2366	−1.58
Non-canonical	**−4.0470**	**0.9693**	**−4.18*****	**−2.8491**	**1.0373**	**−2.75****
Long-distance	0.3020	0.4270	0.71	−1.6934	0.9525	−1.78
WM	**1.0254**	**0.3192**	**3.21****	−0.0128	0.2621	−0.05
IC	−0.4629	0.2434	−1.90	0.2745	0.1978	1.39
RRTs:Non-canonical	−4.8306	2.8773	−1.68	−6.4040	3.2720	−1.96
RRTs:Long-distance	**−2.4789**	**0.9965**	**−2.49***	−4.4416	3.2103	−1.38
RRTs:WM	0.2061	0.7925	0.26	**−1.4354**	**0.6858**	**−2.09***
RRTs:IC	0.2180	0.5113	0.43	−0.2917	0.5447	−0.54
Non-canonical:WM	−1.0678	0.9019	−1.18	0.2972	0.7161	0.41
Non-canonical:IC	0.8730	0.6349	1.37	−0.6680	0.5952	−1.12
Long-distance:WM	0.6645	0.5485	1.21	−1.0441	0.6925	−1.51
Long-distance:IC	−0.0903	0.4484	−0.20	−0.4381	0.4846	−0.90
RRTs:Non-canonical:WM	−4.1062	2.7052	−1.52	−1.2478	1.7417	−0.72
RRTs:Non-canonical:IC	**3.3307**	**1.5653**	**2.13***	−2.5045	1.5498	−1.62
RRTs:Long-distance:WM	0.3966	1.4733	0.27	−1.4311	1.8368	−0.78
RRTs:Long-distance:IC	−1.0064	1.1669	−0.86	−0.6593	1.3478	−0.49

***$p < .001$;
**$p < .01$;
*$p < .05$

The comprehension of *wh*-questions and passives in German children and adolescents with Down syndrome

Eva Wimmer and Martina Penke
TU Dortmund University / University of Cologne

Children with Down syndrome (DS) have difficulty comprehending complex syntactic structures. Ring and Clahsen (2005) attribute these problems to a specific syntactic deficit that affects A(=argument) movement. We investigated the comprehension of passive sentences (A-movement) and *wh*-questions (*wh*-movement) in 22 German children and adolescents with DS compared to mental-age matched typically developing children. The results suggest that an impairment in the syntactic development affecting also *wh*-movement structures occurs in a substantial proportion of children with DS and exceeds their general cognitive disabilities. The error patterns of both tests reveal problems with non-canonical clauses and the use of the Agent-first strategy as in early stages of typical acquisition. Moreover, limitations in phonological working memory might contribute to syntactic processing problems in individuals with DS.

Keywords: Down syndrome, developmental disorders, language impairment, language comprehension, *wh*-questions, passive sentences, sentence-comprehension deficits

1. Introduction: Sentence comprehension problems in individuals with Down syndrome

Down syndrome (DS) is a genetic developmental disorder leading in most cases to a mild to moderate intellectual disability. The behavioural phenotype of DS also encompasses an impaired language development: children and adolescents with DS reach developmental language milestones only with a considerable delay (see Abbeduto, Warren, & Conners, 2007 and Roberts, Chapman, Martin, & Moskowitz, 2008 for overview). In particular syntax, which appears to be an area of "special challenge" for individuals with DS (Abbeduto et al. 2007, p. 253) is often

https://doi.org/10.1075/lald.64.12wim

described as disproportionately affected compared to other verbal domains, such as for instance the lexicon (Fowler, 1990; Thordardottir, Chapman, & Wagner, 2002; Sanoudaki & Varlokosta, 2015). Fowler (1990, p. 322) also underlines that only a few children will acquire complex syntax while the great majority will not.

Problems in understanding syntactic constructions have been reported in children, adolescents, and adults with DS (see Aktas, 2004; Bridges & Smith, 1984; Eriks-Brophy, Goodluck & Stojanovic, 2004; Joffe & Varlokosta, 2007; Perovic, 2006; Ring & Clahsen, 2005; Sanoudaki & Varkokosta, 2015; Witecy & Penke, 2017). In particular, the comprehension of passive sentences has been found to pose severe problems to individuals with DS. Ring & Clahsen (2005) conducted the TAPS (Test of Active and Passive Sentences, van der Lely, 1996) with eight English speaking adolescents with DS (chronological age (CA) 12 to 13 years, mental age (MA) 5–6 years). The TAPS contains passive sentences with action verbs and animate arguments, such as *The fish is eaten by the man*. Subjects are required to choose, out of four pictures, the one matching the sentence they previously heard. The three distractor pictures include a picture showing the action with reversed thematic roles (a fish eating the man), an adjectival interpretation or stative passive (a picture showing an eaten fish on a plate) and a semantic distractor. Ring and Clahsen found the comprehension of passive sentences markedly impaired in their subjects. On average, only about half of the sentences in each test condition including full and short passives were answered correctly (see Ring & Clahsen, 2005, Table 5, p. 487). When pointing to an incorrect picture, subjects with DS usually chose the role reversal distractor. Similar problems with passive sentences in individuals with DS were also observed in other studies, most of them using the TAPS (Bridges & Smith, 1984, Eriks-Brophy et al., 2004; Miolo, Chapman, & Sindberg, 2005; Rubin, 2006; Joffe & Varlokosta, 2007). While there are also reports that some individuals with DS master the passive, these cases are considered to be exceptional (Rubin, 2006) whereas for most individuals with DS a "genuine difficulty with the passive structure" has been reported (Rubin, 2006, p. 94).

To account for the syntactic problems in subjects with DS, several researchers have proposed a domain-specific syntactic deficit (Fowler, 1990; Ring & Clahsen, 2005; Perovic, 2006) affecting specific structural aspects of language independently from the general cognitive impairment observed in individuals with DS. Most prominent is the proposal by Ring and Clahsen (2005), who suggested a deficit in the formation of particular syntactic chains in subjects with DS. This deficit is said to affect A(=argument)-movement which occurs e.g. in the tested passive structures (see Section 1.2) and in the binding of reflexive pronouns and is assumed to be specific for DS.

Whereas Ring and Clahsen assume a very restricted syntactic deficit that only affects A-dependencies such as the passive or reflexive binding, other researchers have argued for a broader syntactic deficit, affecting other constructions and operations as well. A focus of such studies has been on A'-dependencies, such as in *wh*-questions (see Section 1.2). Joffe and Varlokosta (2007) conducted a study with 10 English-speaking children with DS (CA: 5–14 years, MA 3–6 years). They investigated the comprehension of passives and *wh*-questions (*who*- and *which*-questions, subject and object questions) and found significant differences to MA-matched typically developing children (TD) for both these structures in an acting-out-scenario with puppets. They observed that even simple canonical *wh*-questions (e.g. *Who punched the lion?*) led to comprehension problems in the tested subjects with DS. Overall, only about 43% of the pointings were to the correct candidate. A production study by Tsakiridou (2006) with four Greek speaking adolescents with DS (CA: 20–28 years; MA: 7–8 years) elicited *wh*-subject and *wh*-object questions by means of a game with puppets (e.g. *Who did the hippo kiss*?). Here also, a low performance on all question types was reported for every subject, differing significantly from a TD group with a MA of 4 years who had acquired these structures. The studies by Joffe & Varlokosta, as well as Tsakiridou (2006), indicate that not only A-dependencies, but also structures involving A'-dependencies are affected in individuals with DS.

Whereas these studies hint at a syntactic deficit with movement operations, other accounts suggest that the linguistic impairments observed in individuals with DS are due to general or specific cognitive deficits characterizing this syndrome. In order to test this idea, individuals with DS are compared to TD children with a comparable nonverbal cognitive development (i.e. nonverbal MA). If individuals with DS perform similar to control groups of nonverbal MA-matched children when producing or comprehending a tested linguistic structure, their performance is at the level expected by their nonverbal MA. With respect to grammatical comprehension, however, there is conflicting evidence. While some studies have found the performance of children and adolescents with DS in receptive tasks (including passive comprehension tests) similar to TD children of comparable nonverbal MA or comparable verbal comprehension age (Bridges & Smith, 1984; Chapman, Schwartz, & Bird, 1991; Eriks-Brophy et al., 2004), other studies testing syntactic comprehension reported that individuals with DS did not meet nonverbal MA expectations (Abbeduto et al., 2003; Joffe & Varlokosta, 2007; Ring & Clahsen, 2005). Recently, researchers have focused on the role of auditory-verbal short-term memory, or phonological short-term memory, as a potential relevant predictor of language comprehension in individuals with DS (see Laws & Gunn, 2004; Miolo et al., 2005). Weak phonological short-term memory skills are one of the main characteristics of subjects with DS (Baddeley & Jarrold, 2007; Laws & Gunn, 2004 and overview by Naess et al., 2011). Phonological short-term memory is a capacity-limited storage buffer, a subcomponent of the phonological loop that is part of Baddeley's multicomponent model of working memory (e.g. Baddeley, Gathercole, & Papagno, 1998). This buffer is responsible for the temporary storage of phonological material until the information contained in the signal can be further processed. It has been suggested that individuals with DS suffer from a "phonological loop deficit" that particularly affects their capacity to store phonological material in their working memory (Baddeley & Jarrold, 2007, p. 929). As phonological short-term memory skills have been found to be implicated in the comprehension of complex sentences such as passives (see Baddeley, 1990; Vallar & Baddeley, 1984), a deficit in the phonological loop might contribute to the sentence comprehension problems observed in individuals with DS.

1.1 Aim of paper and research questions

The study aims to provide new data on syntactic comprehension in individuals with DS acquiring German, contributing to the debate on whether children and adolescents with DS suffer from a deficit with syntactic movement constructions. Specifically, the study targets the following research questions: (i) Do children with DS display specific problems in comprehending sentences involving A-movement constructions, or do syntactic problems also affect A′-movement constructions?; (ii) Are syntactic problems explicable in terms of the nonverbal cognitive development, i.e. to what extent do children and adolescents with DS perform according to nonverbal MA expectations?; and (iii) Which factors influence syntactic processing in individuals with DS? In particular, we will focus on the role of phonological working memory skills. We investigated passives and *wh*-questions, which are characterized by different syntactic operations (A-movement and A′-movement) in generative syntax.

1.2 Structure and acquisition of German wh-questions and passives

Structure of passives and wh-questions: A-movement and A′-movement
According to standard syntactic theories within the generative framework (Chomsky, 1981), the surface structure of a sentence is established by the movement of constituents out of their base-generated position in an abstract underlying structure. *Wh*-questions and passive clauses can be differentiated at the syntactic level by two different kinds of those movement operations, A-movement (also called NP-movement) and *wh*-movement (or A′-movement) (for a more detailed description see Haegeman, 1994).

In German, as in English, passive clauses (see 1a) involve A-movement, i.e. movement of the internal argument of the sentence from a VP internal position to the subject-position that is an argument (or A-)position. The moved NP and the trace or gap that is coindexed with it form an argument-chain (or A-chain). A-movement, in contrast to *wh*-movement, is case-driven, i.e. the NP that has been assigned the thematic role of patient in the VP has to move to the subject position to receive nominative case in German. The internal argument, i.e. the Agent of the action, can optionally be overtly expressed by a prepositional phrase as in the long passive clauses (1a), or can be left out (short passive clause). Also, a passive differs from a corresponding active sentence (here *Das Mädchen kitzelt den Jungen*, 'The girl is tickling the boy') by a morphological change, as the action verb is realized as a participle while finiteness information is expressed by an auxiliary, either by the verb *werden* (verbal passive reading) or by *be* (which corresponds to an adjectival, static passive reading) (Bartke, 2004; Eisenbeiss, 1993). Typically, the subject is placed before the finite verb resulting in a SVX word order.

Wh-movement takes place for instance in matrix *wh*-questions (see 1b,c). The *wh*-phrase (either the subject or object of the sentence) is base-generated in the VP and is assigned its thematic role by the verb (i.e. Agent or Patient). It is then moved to the leftmost position of the sentence. *Wh*-movement results in a chain (A′-chain). Note that in contrast to English, the word order between the two question types does not differ in German, but the type of question is expressed by the initial *wh*-pronoun (*wer* in subject questions, *wen* in object questions).

(1) a. passive sentence (long)
Der Junge$_i$ wird vom Mädchen $\mathbf{t}_i$ gekitzelt
The-NOM-boy is-being by-DAT the girl tickled
PATIENT (subj) AGENT (prep-obj)
'The boy is being tickled by the girl.'

b. *who*-subject question
Wer$_i$ kitzelt $\mathbf{t}_i$ den Jungen?
who-NOM tickles the-ACC boy
AGENT (subj) PATIENT (obj)
'Who is tickling the boy?'

c. *who*-object question
Wen$_i$ kitzelt der Junge $\mathbf{t}_i$?
who(m)-ACC tickles the-NOM boy
PATIENT (obj) AGENT (subj)
'Who(m) is the boy tickling?'

In German, the unmarked or so-called 'canonical' linear order of arguments is Subject (Agent) before Object (Patient), as in many European languages. Thus, whereas in German subject questions (1b) the described syntactic movement operations lead to the surface constituent order Subject (Agent) – finite Verb – Object (Patient), the resulting structure in object questions as well as in passives deviates from this canonical constituent order: in both these sentence types the patient precedes the Agent: in object questions: Object (Patient) – finite Verb – Subject (Agent), in passives: Subject (Patient) – finite Verb – optional by-Agent, (see 1a,c). A non-canonical word order is generally associated with higher syntactic complexity (see e.g. Caplan, Baker, & Dehaut, 1985), resulting in later acquisition (see below) and larger vulnerability in language disorder syndromes (e.g. Caplan et al., 1985; Grodzinsky, 1995 and Penke, 2015 for overview).

Acquisition of questions and passives in German

German passive sentences can already be produced by young children at the age of 2 or 3 years (Abbot-Smith & Behrens, 2006; Eisenbeiss, 1993). There is some evidence by Eisenbeiss (1993) that children as young as 2 and 3 years can already interpret at least some types of passive sentences (short agentless verbal passives). However, more structured sentence comprehension tests including more items and other types of passive sentences have suggested that the interpretation of passives is mastered only after age 4 (Aschermann, Gülzow, & Wendt, 2004; Bartke, 2004; Schaner-Wolles, Binder, & Tamchina, 1986). Aschermann et al. (2004) tested 60 children from 3 to 6 years in four age-groups and used a sentence-picture matching task as well as an acting-out task. They evaluated the results against chance level and reported a significant jump from 4-year-olds (almost random performance) to 5-year-olds, when the performance yielded a mean accuracy of 87% indicating a performance above-chance level.

The comprehension as well as the production of *wh*-questions is reported to develop between 2 and 4 years of age (see Clahsen, Kursawe, & Penke, 1996; Penner & Kölliker Funk, 1998; Siegmüller, Herzog, & Herrmann, 2005). At approximately 2;06 years, German children realize the syntactic status of the *wh*-pronouns and learn to process these pronouns in sentence-initial position until the age of 3;0 years (see Penner & Kölliker Funk, 1998). There is evidence that the comprehension of some question types is more difficult than others, in particular there is evidence that object questions are more difficult than subject questions (see Roesch & Chondrogianni, 2015; Siegmüller et al., 2005). Our previous research has shown that TD children show three different developmental stages in the comprehension of *wh*-subject and object questions (Wimmer, Rothweiler, & Penke, 2017). At the first stage, both question types cannot be interpreted correctly, leading to a guessing performance, i.e. reactions at chance level. In the next stage, the child shows a so-called subject–object asymmetry, i.e. good comprehension of subject questions, but impaired comprehension of object questions. We attributed the problem with this more complex question type to problems in assigning the correct thematic role to the sentence-initial *who*-phrase, leading to guessing performance. At the final stage, TD children achieve above-chance performance on both *who*-question types, indicating intact syntactic comprehension.

In sum, this view on the two phenomena under study suggests that German children have particular difficulties with non-canonical sentences during syntactic acquisition and that the interpretation of *wh*-questions precedes the interpretation of passive structures.

2. Method

We investigated children and adolescents with DS on several standardized measures of general grammar comprehension, passive comprehension and nonword repetition (NWR), as well as with an experiment testing the comprehension of *wh*-questions.

2.1 Participants

The study sample encompassed 22 children and adolescents with Down syndrome. In addition, a group of 19 TD children served as a control group for the non-standardized question comprehension test. Table 1 provides the relevant characteristics of the subject groups.

Table 1. Characteristics of test group (DS) and control group (TD)

Group	Female/Male	IQ	CA	MA
DS (*n* = 22)	8/14	*M* 55* (*SD* 12)	*M* 10;08 (*SD* 4;1)	*M* 4;03* (*SD* 1;1)
TD (*n* = 19)	9/10	*M* 105 (*SD* 11)	*M* 4;0 (*SD* 0;7)	*M* 4;04 (*SD* 0;10)

Note: Columns 3–5 provide means (M) and standard deviations (SD) in parentheses; IQ: corrected for Flynn effect.
*For one child with DS no IQ score could be determined, MA was estimated as at least 8;0 years (performance exceeds oldest norming sample of IQ screening).

TD Children and children with DS passed a subscale of a standardized nonverbal intelligence screening (reasoning scale of SON-R 2.5–7, Tellegen, Laros & Petermann, 2007; reported reliability: 0.83) to determine their nonverbal cognitive development. TD children displayed an unimpaired cognitive development according to the IQ-screening (IQ above 85), and an age-appropriate language development according to a standardized screening test (SSV, preschool language screening, Grimm, 2003). They were matched to the children with DS with respect to nonverbal MA equivalents calculated from the raw scores of the IQ-screening.

Statistical comparisons revealed no difference between the two subject groups with respect to nonverbal MA that was on average about 4 years (*independent samples t-test*: t (39) = −.273, p = .787, ns). A further inclusion criterion for all tested children was that they had to be monolingual German, had normal hearing and normal or corrected vision. Apart from the cognitive impairment of the subjects with DS, none of the children had additional physical or psychological problems.

2.2 Measures

2.2.1 *Standardized passive comprehension test (TSVK)*

The comprehension of passive sentences was assessed by subtest 4 of the German TSVK test, a standardized test on sentence comprehension in German children (Siegmüller, Kauschke, van Minnen, & Bittner, 2011). For the passive subtest, T-scores are available with respect to six different age groups from 3 to 8 years, based on a normative sample size of 297 monolingual TD children. The subtest consists of a sentence–picture-matching task where children select from an array of three pictures the one matching with the sentence auditorily presented to them. The selection includes the target picture (e.g. a horse kissing a cow), in which the thematic roles of the depicted action match the sentence, a distractor picture with reversed thematic roles (the cow kissing the horse), and another distractor showing the situation after the action has been completed (the kissed cow, i.e. a static or adjectival reading). The material consists of 16 test sentences with passive structures and four active sentences used as fillers (2c). Half of the sentences include long passives entailing the prepositional object (by-Agent), the other half are short passive structures without the by-Agent (see 2a,b).

(2) a. Die Kuh wird vom Pferd geküsst. (long passive)
The-NOM cow is-being by.the-DAT horse kissed
'The cow is being kissed by the horse.'

b. Tim wird geküsst. (short passive)
Tim is-being kissed
'Tim is being kissed.'

c. Tom füttert den Papa. (active filler)
Tom is feeding the-ACC dad.
'Tom is feeding his dad.'

Half of the passive sentences are reversible sentences, allowing for role reversals and an Agent-first interpretation (as in (2)). The other half of the sentences can be considered semantically irreversible as they include inanimate patients, facilitating the target passive reading of the underlying action (e.g. a strawberry picked by a girl). Also, two of the four active sentences depict improbable events, such as an apple picking a man. The position of the target pictures in the page is balanced (top, center or bottom of the page).

2.2.2 *Comprehension of wh-questions*

To test the comprehension of *wh*-questions we used a picture-pointing task that was developed by our research group and has been successfully applied to investigate the comprehension of *wh*-questions in TD and developmentally impaired German-speaking children (see Wimmer et al., 2017). The pictures depict semantically reversible actions performed by a boy and a girl (e.g. feeding, see examples in Figure 1).

Das Mädchen hat Hunger. ('The girl is hungry').
Der Junge füttert das Mädchen. ('The boy is feeding the girl')
Das schmeckt lecker. ('It tastes delicious')

a. *Wer füttert das Mädchen?* ('Who is feeding the girl?')
b. *Wen füttert der Junge?* ('Who(m) is the boy feeding?')

Figure 1. Example of a test picture and stimuli in the picture-pointing task

The scenario depicted in each picture is introduced by three short spoken sentences describing who is doing what to whom in the picture (see Figure 1). Then a short *wh*-argument question is asked to which the child has to respond by pointing to the corresponding person (left or right) presented in the picture. Only one question, either *who*-subject or *who*-object, is asked per picture. In total, ten *who*-subject questions (*wer*-questions) and ten *who*-object questions (*wen*-questions) are presented. The material is controlled for a number of potentially influencing factors, e.g. with regard to the role of the boy or girl in the picture and the position of the persons in the picture (Agent/Patient; left/right – balanced across conditions) and the kind of verbs used in the questions (transitive lexical verbs that are part of German preschool children's vocabulary). The 20 questions usually consist of four words (see Figure 1a, b) and are presented in two blocks of ten questions (each five subject and object questions in randomized order). No more than two identical question types follow each other during presentation. The task is introduced by a short training phase familiarizing the children with the pointing task and ensuring that they recognize the persons in the pictures as 'boy' and 'girl'. During the test phase children could take their time to answer and the question was repeated once more if requested.

2.2.3 *Background measures: General grammar comprehension and phonological working memory*

The TROG-D (Fox, 2011, German adaption of the *Test for Reception of Grammar*) was conducted with 19 of the tested 22 children with DS. It is a standardized tool to investigate verbal comprehension of different grammatical structures with increasing complexity (Cronbach's $\alpha = .090$) and it provides norming data for German children aged 3 to 10 years. The participants were instructed to identify the picture matching a word or sentence spoken by the experimenter out of an array of four pictures. The test consists of 21 blocks of four items each and includes a stop criterion that is reached if the participant produces at least one incorrect reaction in five consecutive blocks. In this study, the TROG-D is used as a background measure testing to what extent the two sentence comprehension tests of this study are related to a general test on grammar comprehension.

A nonword repetition (NWR) test (Grimm, 2003) served as a measure of phonological working memory skills. Norming data are available for German children aged 3 to 5 years. The test consists of 18 nonwords of two to five syllables that have to be repeated as accurately as possible by the participants.

2.3 Data analysis

For all three standardized tests (TROG-D, NWR, passive test) we determined raw scores for each participant with DS according to the test manuals. In the TROG-D we counted the number of correct blocks (a block is scored correct if all four items are correct; maximum raw score: 21). In the NWR-test the raw scores correspond to the absolute number of correctly repeated nonwords (maximum score: 18) and in the passive test raw scores are determined by the number of correct pointings to the target picture (maximum score: 16). To evaluate if a child displays an age-equivalent performance (with CA and nonverbal MA considered separately), the individual raw scores obtained in these tests were transformed to T-scores based on the CA or MA of the child or adolescent. Participants scoring below the critical T-value (T = 40) are considered to display a substandard performance, i.e. a deficit.[1]

The passive test and the question comprehension task were subjected to further analyses. With respect to the passive test, we calculated mean accuracy scores for the 16 passive test sentences as well as for the four active sentences. For the

1. Note that in some cases no exact T-score could be determined for the three standardized tests if CA or MA exceeds the oldest age group of the norming sample in these tests. However, a substandard performance was revealed if the raw scores fell *below* the score required by the oldest norming group of the test.

passive sentences, we also determined the mean frequency of picking the two distractor pictures. *Paired-samples t-tests* were used to compare different types of passive sentences and to compare passives with active sentences. To determine how many participants already master the passive, we analyzed the individual performance scores with respect to chance level as an indicator of guessing. Therefore, the data were compared to a binomial distribution with an alpha level of 0.05. With 16 analyzable sentences, a subject's performance is not different from chance (being one third with three options to react) if three to eight sentences are correctly understood. A correct response of 9 to 16 pictures is indicative of above-chance performance, whereas correct responses to two or less pictures suggest below-chance performance.

For the question comprehension test, mean accuracy scores of the analyzable reactions were calculated, i.e. the number of unambiguous pointings (or in rare cases verbal answers) to the person (boy or girl) in the picture. A two-factorial mixed analysis of variance (ANOVA) was conducted to quantify differences between groups (factor GROUP: DS, TD) and differences with respect to the *wh*-question type (within-subjects factor QUESTION TYPE: *who*-subject question, *who*-object question). To closer analyze individual performance patterns, we determined how many subjects showed intact syntactic comprehension, defined as above-chance interpretation of the non-canonical object questions. In this task, a subject's performance is not different from chance (50%, two options to react) if three to seven pointing gestures out of 10 in the object question condition are correct (above-chance: 8 to 10, below-chance: 0–2).

For both sentence comprehension tests (passives and *wh*-questions), we considered effects of potentially influencing factors for the group with DS such as cognitive development (as determined by nonverbal MA), maturational factors (indexed by CA) and other environmental factors (such as socioeconomic background) (see Fowler, 1990; Thordardottir et al., 2002) by computing correlational analyses (*Pearson's r correlation*) between those variables and overall accuracy scores in these tasks (see Section 3.4.). Socioeconomic status (SES) was based on the level of education of the child's mother on a 9-level scale (according to the International Standard Classification of Education 2011, see UNESCO, 2012). The influence of phonological working memory skills on sentence comprehension was tested by correlational analyses between the children's results in the NWR test and overall results in the two sentence comprehension tests. We also correlated the raw sores of TROG-D with the accuracy scores of the two sentence comprehension tests to measure the interaction between general grammar comprehension skills and more specific syntactic comprehension abilities.

3. Results

3.1 Results of the passive test (TSVK)

Figure 2 shows the reactions of children with DS in the standardized TSVK passive test for the 16 passive sentences. In only 53% of the cases (*M* = 8.4 out of 16, *SD* = 3.3, range = 3–15) children with DS pointed to the target picture. In 40% of the cases the picture depicting the action with reversed thematic roles was chosen, in the remaining few cases children pointed to the picture showing the completed action. That is, in the majority of incorrect cases (83.2%) children preferred the interpretation of the subject NP as the Agent of the action instead of showing an adjectival or static passive reading of the clause.

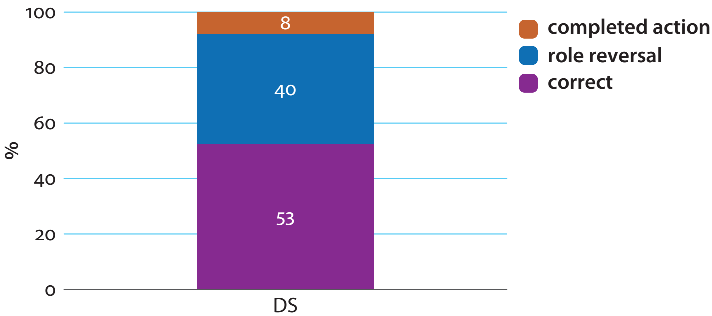

Figure 2. Pointing reactions with passives of children with DS in the TSVK passive test

No significant differences were found with respect to the type of the passive clause (short vs. long passives) or with respect to the improbability or reversibility of the action (reversible vs. irreversible clauses) (*paired samples t-test*: type of passive: $t(21) = 1.262$, $p = .221$, ns; reversibility: $t(21) = -.086262$, $p = .932$, ns). The comprehension performance for the active filler sentences was significantly better than for passive sentences (accuracy scores actives vs. passives: $t(21) = 3.091$, $p = .006$). Children with DS pointed in 70.5% (*M* = 2.8 out of 4, *SD* = 1) of the cases to the target picture.

With respect to their CA, all but two subjects (the youngest ones: CA 4;6 and 5;5) had T-scores indicating a substandard performance in passive comprehension according to the TSVK. However, with the exception of one subject whose nonverbal MA exceeds the norms of the TSVK, all subjects with DS scored within the norms of the TSVK with respect to their nonverbal MA. This does not mean that passive comprehension is mastered in these individuals with DS. Evaluating the performance in the passive test against chance-level, only nine subjects (41%) displayed an above-chance pattern on passive sentences. The majority of 13 individuals with DS (59%) displayed a performance at chance level indicating a guessing performance as response to passive sentences. This observation strongly indicates that these latter subjects have not yet acquired the passive.

3.2. Results of the who-question comprehension task

Figure 3 displays the results for the children with DS and the TD children with respect to the mean accuracy scores achieved in the question comprehension test, overall and with respect to the two question types tested. For children and adolescents with DS, on average, only 71.8% (*SD* = 29.1%, range 5.0–100%) of the reactions were correct compared to 92.9% (*SD* = 10.8%, range 60.0–100%) of the reactions in the control group. In both *wh*-question conditions, children with DS performed worse than TD children despite a similar nonverbal MA. This is supported by a significant main effect of GROUP in the ANOVA ($F(1,39) = 8.857$, $p < .005$, Cohen's $d = .95$).

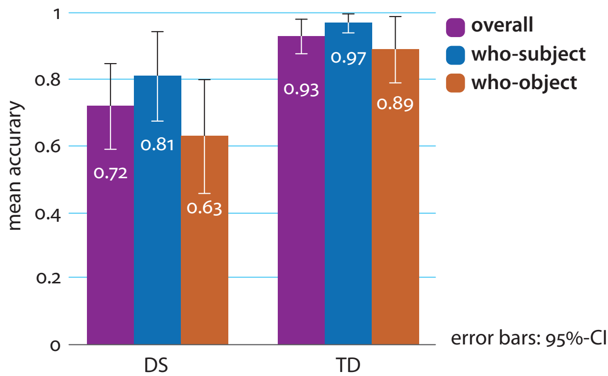

Figure 3. Mean accuracy scores of children with DS and TD in the question comprehension task

Figure 3 also illustrates that *who*-subject questions were better understood than *who*-object questions in both groups of children. The analysis yields a significant main effect of QUESTION TYPE ($F(1,39) = 6.997$, $p < .012$, Cohen's $d = .85$). Although more pronounced for the children with DS, this effect is independent from the group of children (no significant interaction between GROUP and QUESTION TYPE ($F(1,39) = 1.089$, $p = .303$, ns, Cohen's $d = .33$).

With respect to the performance of the TD children, only three of the younger participants (MA from 3;4 to 3;8) showed problems in comprehending *who*-questions and performed at or below chance with non-canonical object questions. The great majority of TD children (84.2%) had no problems in interpreting *who*-questions and achieved above chance performance for both sentence types. In contrast, only 13 (59.1%) of the 22 participants with DS performed above chance on the non-canonical *wen*-questions, whereas the remaining nine subjects had problems in understanding this question type. Three of these nine individuals displayed a behavior at chance level, suggesting a guessing strategy with the tested object questions. For the other six subjects, the performance for *who*-object questions was *below*-chance level. In all but one of these participants, the below-chance performance seems to follow from an Agent-first strategy throughout the experiment, i.e. these five subjects consistently assigned the Agent role to the first NP constituent of the sentence independent of the sentence type. In *who*-object questions, but not in *who*-subject questions, such a strategy leads to the wrong assignment of thematic roles, i.e. the child incorrectly points to the Agent in the picture instead of the Patient. The remaining child seemed to apply a strategy of consistently pointing to the last heard constituent (i.e. *the girl* in the sentence: *Who is feeding the girl*?) which always led to the incorrect response (only one correct reaction overall). In contrast to TD children, problems on comprehending *wh*-object questions even occurred at a MA above 3;08 years and even in the mentally oldest subject (MA > 7;11 years).

3.3 Results of background measures: TROG-D and NWR

In the TROG-D, the raw scores depicting the number of correct blocks ranged between 2 and 15 for the 19 participants with DS tested ($M = 5.2$, $SD = 3.4$, range = 2–15). The T-scores of all but two subjects were within the range of the norming sample with respect to their nonverbal MA. With respect to their CA, however, none of the subjects scored within the age norm.

In the NWR task measuring phonological working memory, the 21 children with DS achieved a mean raw score of 4.3 ($SD = 3.7$). Only nine of them were within the normal range for their nonverbal MA, whereas the remaining majority performed below the critical T-score of 40. Relative to CA, only one subject scored clearly within the age norm. Four subjects had raw scores that were adequate (i.e. T-scores 41–55) with respect to the oldest norm group of this test (5;0–5;11 years), but appear deviant compared to their high CA (8;06–14;07 years). The scores of the remaining 17 participants fell below the age norm based on their CA.

3.4 Comparison of tests and correlational analyses

Comparing both syntactic tests within the group of DS subjects, the accuracy scores achieved with the non-canonical structures in these tests, i.e. *who*-object questions and passive sentences, correlated positively with one another ($r(20) = 502$, $p = .017$). Eight of the nine participants who comprehended passive clauses above chance could also interpret *wh*-object questions as well as *wh*-subject questions target-like (see Figure 4), i.e. with accuracy scores at ceiling (overall accuracy: $M = 98.1\%$). Eight participants of the group could neither interpret questions nor passives above chance. Remarkably, there were five subjects with DS who understood object questions well, but failed with passive clauses. Thus, it seems that the acquisition of short *wh*-questions precedes the acquisition of passives.

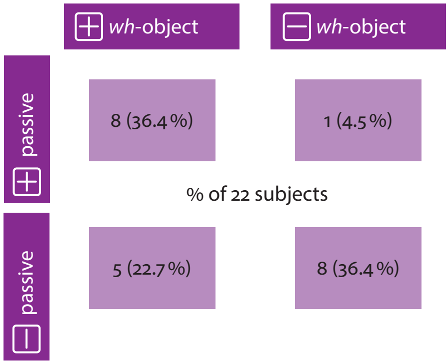

Figure 4. Patterns for *wh*-questions and passives (+ above chance; – chance/below chance)

Only one subject performed above-chance on passive sentences (68.8% accuracy) in the presence of a deficit with object questions (see Figure 4). A consistent Agent-first strategy in the question task and the dominant error type of role reversals in the passive task indicate that neither *wh*-questions nor passives sentences are fully acquired by this child.

The correlational analyses between the overall results achieved by the 22 subjects with DS in the two sentence comprehension tests, *wh*-questions and passives, and potentially influencing factors revealed that only one of those factors yielded a significant interrelation, namely phonological working memory skills (see Table 2). However, the relationship between syntactic skills and phonological working memory is not without exceptions: among the "low" performers (i.e. those children that displayed impaired comprehension with *wh*-questions as well as passives), there are also two children with phonological working memory skills within the norm of TD peers (T-score of 40 and 53, based on nonverbal MA). Similarly, two of the high performers in passive and question comprehension performed below the critical T-score (each T-score 39). This indicates that in some cases, syntactic comprehension is impaired despite MA-equivalent phonological working memory skills and vice versa.

Table 2 moreover shows that the overall accuracy scores of both sentence comprehension tests (*wh*-questions and passive test) for children and adolescents with DS are positively correlated with the raw scores achieved in the standardized

Table 2. Results of correlational analysis in question comprehension and TSVK (passive) test

Pearson's r correlations of variables with test results (overall accuracy)	*wh-questions*	*passive test*
CA: chronological age	.276	.340
MA: nonverbal mental age	–.014	.347
Nonverbal IQ	–.019	.335
Phonological working memory	**.526***	**.801*****
SES: socioeconomic status (i.e. education level of child's mother, $N = 20$)	.064	.388
TROG-D	**.532***	**.794*****

Note: *p <.05, *p < .01, ***p < .001, significant results and large effects in bold

TROG-D testing general grammar comprehension. Remarkably, of those eight children who showed impaired syntactic comprehension (for non-canonical questions and for passives) all but two children achieved a performance *within* mental-age equivalent norms in the TROG-D. Only for two of those children, the standardized TROG-D indicated a substandard performance.

4. Discussion

The aim of the study was to investigate the comprehension of two constructions involving A- and A′-dependencies in German individuals with DS. The results are summarized and discussed as follows in light of the central research questions.

i. *Individuals with DS display severe difficulties in the comprehension of structures involving A-movement (passives) as well as A′-movement (wh-questions). Syntactic problems predominantly affect non-canonical sentence types.*

Of the 22 tested children and adolescents with DS, only about one third (eight subjects) performed above chance on passive clauses as well as on *wh*-questions. The other 14 children had severe comprehension difficulties related to at least one of these sentence types with passive sentences more affected than questions. The data indicate, for one, that structures involving A-movement (passives) can be affected in German children and adolescents with DS, a finding that corroborates the observations made in several studies, as e.g. by Ring and Clahsen (2006) or Rubin (2006). Secondly, the question comprehension test revealed that question structures involving A′-dependencies can also be severely impaired in DS, which is in line with previous studies observing deficits in the comprehension and/or production of *wh*-questions in these individuals (Joffe & Varlokosta, 2007; Tsakiridou, 2006). We can therefore conclude that a supposed syntactic deficit in individuals with DS is not specifically related to an inability to form A-chains, as has been suggested by Ring and Clahsen (2005) and Perovic (2006).

A closer analysis of the result patterns of both tests suggests that the observed difficulties are rather determined by the complexity of the sentence structure than by the type of syntactic movement. Canonical, Subject (Agent) initial sentences (subject questions and simple actives) were significantly better comprehended than non-canonical sentences (object questions and passives) with the object or the Patient argument in sentence-initial position. Individually, this was reflected by above-chance patterns with canonical sentences and chance/below chance-patterns with non-canonical clauses. This asymmetric pattern, often referred to as 'subject–object asymmetry' (see Section 1.2), is a typical performance pattern in individuals with language disorders that has also been attested in individuals with DS with respect to active vs. passive sentences (Bridges & Smith, 1985; Ring & Clahsen, 2006; Rubin, 2006). However, whereas the asymmetry is typically associated with a guessing strategy with the non-canonical sentence type (evidenced by chance-performance), we observed that individuals with DS often rely on the use of an Agent-first strategy leading to below-chance-performance, a dominant error pattern in the question task. Independent of question type, the *wh*-pronoun was interpreted as referring to the actor of the action. Although we did not find evidence of a consistent use of the Agent-first strategy in the passive task, the reactions of the low performers with DS in this task (although at chance-level) might not reflect a simple guessing strategy. Guessing should have led to an equal choice between the two incorrect options. However, it was usually the distractor picture with reversed thematic roles that was incorrectly chosen and not the one reflecting an adjectival passive reading, similar to the observations in the studies by Ring & Clahsen (2005) and Rubin (2006). Thus, from the error pattern observed in both the question comprehension and the passive task, it is evident that subjects with DS often rely on the use of a basic heuristic strategy, the Agent-first strategy.[2] This does not only show that syntactic comprehension is deficient, but that the affected children might not be able to distinguish between different sentence types yet which reflects a very early stage in language acquisition (see Bever, 1970 for the use of the Agent-first strategy in early acquisition).

2. Note that an Agent-first strategy has been observed in aphasic speakers (Penke, 2015) and has also been described as a general parsing principle in adults that initially assigns the subject role to the sentence-initial argument, e.g. in the case of an ambiguous *wh*-phrase (see Schlesewsky, Fanselow, Kliegl & Krems, 2000).

The comparison of the individual results of both syntactic tests further revealed that in some children the comprehension of non-canonical *wh*-questions is already mastered whereas the comprehension of passive sentences is still problematic. Individuals with DS therefore seem to adhere to the typical developmental course with the comprehension of questions preceding comprehension of passive clauses. However, due to the use of cross-sectional data, it remains unclear if the passive will be mastered by those children/adolescents with DS who succeeded in interpreting questions. Moreover, in light of the fact that the questions were very short and only encompassed *who*-subject and *who*-object-questions, it could be the case that other, more complex *wh*-questions (e.g. *which*-questions) or other sentence structures that are already mastered by TD children of similar MA could pose a problem for those participants with DS. This issue has, however, to be left to further studies.

We also found that individuals with DS are no homogenous group with respect to their ability to comprehend passives and *wh*-questions. The data demonstrate that a subgroup of children with DS is apparently able to acquire complex sentence structures – such as passives and non-canonical *wh*-questions. The observation that nine children of a study sample of 22 children showed above-chance performance with passives contradicts the assumption that individuals with DS cannot produce or comprehend syntactically complex sentences (e.g. Fowler, 1990; Rubin, 2006). We therefore assume that children/adolescents with DS, in principle, are able to acquire complex syntactic constructions which also concurs with similar claims by work from Eriks-Brophy et al. (2004) and Thordardottir et al. (2002) who based this assumption on the analysis of narrative samples.

ii. *For some individuals with DS, the impairment goes beyond the cognitive disability*.

In the TROG-D most of the subjects performed within MA-equivalent norms. This supports previous studies that have also found general grammatical comprehension skills of children with DS to commensurate with MA (e.g. Bridges & Smith, 1984; Chapman et al., 1991). As expected, comprehension of passives and *wh*-questions correlated with the raw scores obtained in the TROG-D. However, whereas in the TROG-D most of the subjects performed within MA-equivalent norms, a substantial proportion of 40.9% (9 children) deviated from MA-matched controls in a finer-grained comprehension test on *wh*-questions. Contrary to TD control children of comparable nonverbal MA, these individuals with DS could not interpret non-canonical *who*-object questions, indicating impaired syntactic comprehension that goes beyond the attested intellectual disability. This finding cannot be accounted for by the nonverbal cognitive development of the tested individuals with DS as the performance with respect to *who*-object questions indicates a developmental stage that is overcome by TD children at a comparable cognitive developmental age. Also, we failed to find a significant interrelation between nonverbal MA and the results of the two sentence comprehension tests. The conclusion that the difficulties with complex sentences in our test group cannot be accounted for by their general cognitive development (nonverbal MA, nonverbal IQ) is in line with the studies by Perovic (2006) and Ring & Clahsen (2005).

iii. *Limitations in phonological working memory might contribute to syntactic comprehension problems in individuals with DS.*

While the problems with complex sentence structures observed in children and adolescents with DS in this study might be interpreted in favor of broader syntactic deficit accounts assuming difficulties with particular movement operations or syntactic chains, they might not result from a deficit of syntax proper. An alternative approach would be to seek the deficit in more specific cognitive impairments, such as limitations of processing capacities that are said to be involved in comprehending complex syntactic structures. With respect to DS, Baddeley & Jarrold (2007) suggested a specific deficit with the phonological storage function within working memory. In this study, we did not only find that phonological working memory, as measured by a NWR task, was severely impaired and not MA-equivalent in most subjects, but we also found positive correlations between this test and the results achieved in both sentence comprehension tests. This observation is in accordance with studies that also reported performance on sentence comprehension tasks to be related to auditory-verbal memory tests in children and adolescents with DS (Chapman et al., 2002; Laws & Gunn, 2004; Miolo et al., 2005). These findings suggest that verbal short-term memory capacities might play a role for the comprehension of syntactically complex sentences in individuals with DS.

Specifically, a supposed deficit with the phonological storage buffer (Baddeley & Jarrold, 2007) might affect syntactic movement operations where the moved constituent (for instance a *who*-object phrase) or its features have to be kept in memory until its trace in the base position is encountered and the thematic role of the moved constituent can be identified and processed. Thus, children with DS are likely to fail in parsing complex structures that require the processing of non-canonical word ordering, long syntactic dependencies, or of more syntactic operations. However, a closer investigation of the cognitive processes underlying tasks such as NWR is necessary to establish how specific operations of phonological working memory interact with different syntactic constructions respectively operations. Findings from such an investigation could clarify why subjects can comprehend object questions, but not passives (see Figure 4) although both these non-canonical structures are syntactically complex.

Moreover, as there were also children who could not even interpret short questions (neither passives) in spite of MA-equivalent phonological working memory skills and vice versa, a word of caution is required to claim a *causal* relationship between deficits in syntactic comprehension and deficits in phonological working memory. Also, factors other than grammatical complexity might additionally tax phonological working memory, such as the number of pictures presented in a sentence-picture matching task). As to date only few studies (such as the work by Miolo et al., 2005) directly related performance of tasks measuring phonological working memory to grammatical comprehension performance in individuals with DS, further studies are needed to explore this relationship more closely.

5. Conclusion

This study contributes to our understanding of the syntactic abilities of individuals with DS and the debates about the basis of the deficit underlying syntactic problems in DS. Our data reveal that a substantial subgroup of German children and adolescents with DS had severe problems in comprehending short *wh*-questions and/or passives. Contrary to the assumption that a specific deficit in forming A-movement chains characterizes DS, these data indicate that structures involving A′-movement are affected as well. The data, moreover, suggest that the syntactic comprehension problems of some individuals with DS cannot simply be accounted for by their general cognitive development. However, this study also demonstrates that some children and adolescents with DS are indeed able to acquire complex structures such as non-canonical *wh*-questions and passives. Correlational analyses indicated that neither CA, nonverbal MA, nonverbal IQ, nor SES had an impact on the outcome in both tasks. Only phonological working memory skills had a significant effect on the comprehension performance. However, a closer investigation of the role of phonological working memory skills in comprehending particular sentence structures is required to clarify why some complex structures (such as object questions), are easier to interpret than others (such as passives). There is also a need for longitudinal data that could shed more light on the issue if the observed deficits are persistent, i.e. if a plateau is reached at some point during the syntactic development of individuals with DS.

Acknowledgements

We are grateful to parents and participants, schools and intervention centers for their cooperation.

Funding

The work reported in this paper was supported by a German research foundation (DFG) grant to Eva Wimmer and Martina Penke (grant no. 4130/2-1 and 683/3-1).

References

Abbeduto, L., Murphy, M. M., Cawthon, S. W., Richmond, E. K., Weissman, M. D., Karadottir, S., & O'Brien, A. (2003). Receptive language skills of adolescents and young adults with Down syndrome or fragile X syndrome. *American Journal on Mental Retardation*, 108(3), 149–160. https://doi.org/10.1352/0895-8017(2003)108<0149:RLSOAA>2.0.CO;2

Abbeduto, L., Warren, S. F., & Conners, F. A. (2007). Language development in Down syndrome: From the prelinguistic period to the acquisition of literacy. *Mental Retardation and Developmental Disabilities Research Reviews*, 13(3), 247–261. https://doi.org/10.1002/mrdd.20158

Abbot-Smith, K., & Behrens, H. (2006). How known constructions influence the acquisition of other constructions: The German passive and future constructions. *Cognitive Science*, 30(6), 995–1026. https://doi.org/10.1207/s15516709cog0000_61

Aktas, M. (2004). *Sprachentwicklungsdiagnostik bei Kindern mit Down- Syndrom. Entwicklung eines diagnostischen Leitfadens zum theoriegeleiteten Einsatz standardisierter Verfahren* (Unpublished doctoral dissertation). University of Bielefeld. Retrieved from <https://pub.uni-bielefeld.de/publication/2302157> (4 May, 2020).

Aschermann, E., Gülzow, I., & Wendt, D. (2004). Differences in the comprehension of passive voice in German-and English-speaking children. *Swiss Journal of Psychology*, 63(4), 235–245. https://doi.org/10.1024/1421-0185.63.4.235

Baddeley, A. (1990). *Human memory*. Mahwah, NJ: Lawrence Erlbaum Associates.

Baddeley, A., & Jarrold, C. (2007). Working memory and Down syndrome. *Journal of Intellectual Disability Research*, 51(Pt 12), 925–931. https://doi.org/10.1111/j.1365-2788.2007.00979.x

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Bartke, S. (2004). Passives in German children with Williams syndrome. In S. Bartke & J. Siegmüller (Eds.), *Williams syndrome across languages* (pp. 345–370). Amsterdam: John Benjamins. https://doi.org/10.1075/lald.36.19bar

Bever, T. G. (1970). The cognitive basis for linguistic structures. In R. Hayes (Ed.), *Cognition and language development* (pp. 279–362). New York, NY: Wiley & Sons.

Bridges, A., & Smith, J. V. (1984). Syntactic comprehension in Down's syndrome children. *British Journal of Psychology*, 75(2), 187–196. https://doi.org/10.1111/j.2044-8295.1984.tb01890.x

Caplan, D., Baker, C., & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition*, 21, 117–175. https://doi.org/10.1016/0010-0277(85)90048-4

Chapman, R. S., Hesketh, L. J., & Kistler, D. J. (2002). Predicting longitudinal change in language production and comprehension in individuals with Down syndrome: Hierarchical linear modeling. *Journal of Speech, Language, and Hearing Research*, 45(5), 902–915. https://doi.org/10.1044/1092-4388(2002/073)

Chapman, R. S., Schwartz, S. E., & Bird, E. K. (1991). Language skills of children and adolescents with Down syndrome, I: Comprehension. *Journal of Speech and Hearing Research*, 34(5), 1106–1120. https://doi.org/10.1044/jshr.3405.1106

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Clahsen, H., Kursawe, C., & Penke, M. (1996). Introducing CP: Wh-questions and subordinate clauses in German child language. In C. Koster & F. Wijnen (Eds.), *Proceedings of the Groningen Assembly on Language Acquisition* (pp. 5–22). Groningen: University of Groningen, Center for Language and Cognition.

Eisenbeiss, S. (1993). Auxiliaries and the acquisition of the passive. In E. V. Clark (ed.), *The Proceedings of the 25th Annual Child Language Research Forum* (pp. 235–242). Stanford, CA: CSLI.

Eriks-Brophy, A., Goodluck, H., & Stojanovic (2004). Comprehension and production of syntax in high-functioning individuals with Down syndrome. In M-O. Junker, M. McGinnis, & Y. Roberge (Eds.), *Proceedings of the 2004 annual conference of the Canadian Linguistic Association*. Montreal (Canada).

Fowler, A. E. (1990). Language abilities in children with Down syndrome: Evidence for a specific syntactic delay. In D. Cicchetti & M. Beeghly (Eds.), *Children with Down syndrome* (pp. 302–328). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511581786.010

Fox, A. V. (2011). *TROG-D Test zur Überprüfung des Grammatikverständnisses*. Idstein: Schulz-Kirchner.

Grimm, H. (2003). *Sprachscreening für das Vorschulalter (SSV)* [*Language screening for preschool children*]. Göttingen: Hogrefe.

Grodzinsky, Y. (1995). A restrictive theory of agrammatic comprehension. *Brain and Language*, 50, 27–51. https://doi.org/10.1006/brln.1995.1039

Haegeman, L. (1994). *Introduction to government and binding theory*. Oxford: Blackwell.

Joffe, V., & Varlokosta, S. (2007). Patterns of syntactic development in children with Williams syndrome and Down's syndrome: Evidence from passives and wh-questions. *Clinical Linguistics & Phonetics*, 21(9), 705–727. https://doi.org/10.1080/02699200701541375

Laws, G., & Gunn, D. (2004). Phonological memory as a predictor of language comprehension in Down syndrome: A five-year follow-up study. *Journal of Child Psychology and Psychiatry*, 45(2), 326–337. https://doi.org/10.1111/j.1469-7610.2004.00224.x

Miolo, G., Chapman, R. S., & Sindberg, H. A. (2005). Sentence comprehension in adolescents with Down syndrome and typically developing children: Role of sentence voice, visual context, and auditory-verbal short-term memory. *Journal of Speech, Language, and Hearing Research*, 48(1), 172–188. https://doi.org/10.1044/1092-4388(2005/013)

Næss, K.-A. B., Lyster, S.-A. H., Hulme, C., & Melby-Lervåg, M. (2011). Language and verbal short-term memory skills in children with Down syndrome: A meta-analytic review. *Research in Developmental Disabilities*, 32(6), 2225–2234. https://doi.org/10.1016/j.ridd.2011.05.014

Penke, M. (2015). Syntax and language disorders. In T. Kiss & A. Alexiadou (Eds.), *Syntax – Theory and analysis. An international handbook* (Vol. 3, pp. 1833–1874). Berlin: De Gruyter Mouton.

Penner, Z., & Kölliker Funk, M. (1998). *Therapie und Diagnose von Grammatikerwerbsstörungen: Ein Arbeitsbuch*. Luzern: SZH/SPC.

Perovic, A. (2006). Syntactic deficit in Down syndrome: More evidence for the modular organisation of language. *Lingua*, 116, 1616–1630. https://doi.org/10.1016/j.lingua.2005.05.011

Ring, M., & Clahsen, H. (2005). Distinct patterns of language impairment in Down's syndrome and Williams syndrome: The case of syntactic chains. *Journal of Neurolinguistics*, 18(6), 479–501. https://doi.org/10.1016/j.jneuroling.2005.06.002

Roberts, J. E., Chapman, R. S., Martin, G. E., & Moskowitz, L. (2008). Language of preschool and school-age children with Down syndrome and fragile X syndrome. In J. E. Roberts, R. S. Chapman, & S. F. Warren (Eds.), *Speech and language development and intervention in Down syndrome and fragile X syndrome* (pp. 77–115). Baltimore, MD: Paul H. Brookes.

Roesch, A. D., & Chondrogianni, V. (2015). The use of case in the comprehension of wh-questions in German-speaking children with and without SLI. In C. Hamann & E. Ruigendijk (Eds.), *Language acquisition and development: Proceedings of GALA 2013* (pp. 379–402). Newcastle upon Tyne: Cambridge Scholars.

Rubin, M. (2006). The passive in adolescents with Down syndrome: A case study. *Down's Syndrome, Research and Practice*, 11(2), 88–96. https://doi.org/10.3104/case-studies.319

Sanoudaki, E., & Varlokosta, S. (2015). Pronoun comprehension in individuals with Down syndrome: The role of age. *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, 50(2), 176–186. https://doi.org/10.1111/1460-6984.12122

Schaner-Wolless, C., Binder H., & Tamchina, D. (1986). Frühes Leid mit der Leideform. Zum Passiverwerb des Deutschens. *Wiener Linguistische Gazette*, 37, 5–38.

Schlesewsky, M., Fanselow, G., Kliegl, R., & Krems, J. (2000). The subject preference in the processing of locally ambiguous wh-questions in German. In B. Hemforth & L. Konieczny (Eds.), *German sentence processing* (pp. 65–93). Dordrecht: Springer. https://doi.org/10.1007/978-94-015-9618-3_3

Siegmüller, J., Herzog, C., & Herrmann, H. (2005). Syntaktische und lexikalische Aspekte beim Verstehen von Informationsfragen – Ein Vergleich zwischen Kindern mit spezifischer Sprachentwicklungsstörung und Kindern mit ungestörter Sprachentwicklung. *L.O.G.O.S. Interdisziplinär*, 13, 29–35.

Siegmüller, J., Kauschke, C., van Minnen, S., & Bittner, D. (2011). *Test zum Satzverstehen von Kindern (TSVK): Eine profilorientierte Diagnostik der Syntax*. München: Elsevier.

Tellegen, P. J., Laros, J. A., & Petermann, F. (2007). *SON-R 2½-7. Non-verbaler Intelligenztest* [*SON-R 2½-7. Snijders-Oomen Nonverbal Intelligence Test*] (Revised 2nd ed. with German norms). Göttingen: Hogrefe.

Thordardottir, E. T., Chapman, R. S., & Wagner, L. (2002). Complex sentence production by adolescents with Down syndrome. *Applied Psycholinguistics*, 23(02), 163–183. https://doi.org/10.1017/S0142716402002011

Tsakiridou, M. (2006). The linguistic profile of Down's syndrome subjects: evidence from Wh-movement construction. *SOAS Working Papers in Linguistics*, 14, 227–248.

Vallar, G., & Baddeley, A. (1984). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior*, 23(2), 151–61. https://doi.org/10.1016/S0022-5371(84)90104-X

Van der Lely, H. K. (1996). *The Test of Active and Passive Sentences (TAPS)*. London: DLDCN.

Wimmer, E., Rothweiler, M., & Penke, M. (2017). Acquisition of who-question comprehension in German children with hearing loss. *Journal of Communication Disorders*, 67, 35–48. https://doi.org/10.1016/j.jcomdis.2017.05.003

Witecy, B., & Penke, M. (2017). Language comprehension in children, adolescents, and adults with Down syndrome. *Research in Developmental Disabilities*, 62, 184–196. https://doi.org/10.1016/j.ridd.2017.01.014

Index

The present volume presents research on language processing and language disorders. Topics range across typical language processing, child developmental language disorders, adult neurodegenerative disorders and neurological bases of typical or impaired brains. The chapters cover a number of linguistic phenomena, including relative clauses, empty categories, determiner phrases and inflectional morphology. Work in this collection uses a variety of experimental methods, both online and offline, such as eye tracking, reaction times, Event Related Potentials, picture selection, sentence elicitation and picture matching tasks. This book will be useful for linguists, speech therapists, and psycholinguists working on the processing of morphosyntax.

ISBN 978 90 272 0763 0

9 789027 207630

JOHN BENJAMINS PUBLISHING COMPANY