ISSN 2385-4138 (digital)
https://doi.org/10.5565/rev/isogloss.276

Isogloss 2023, 9(1)/1
1-29

Comprehension of double-center embedded relatives in Italian: a case for hierarchical intervention

Marco Sala
Department of Psychology, University of Milano-Bicocca
m.sala116@campus.unimib.it

Giulia Bettelli
Department of Psychology, University of Milano-Bicocca
g.bettelli@campus.unimib.it

Beatrice Giustolisi
Department of Psychology, University of Milano-Bicocca
beatrice.giustolisi@unimib.it

Alessandra Vergallito
Department of Psychology & NeuroMi, University of Milano-Bicocca
alessandra.vergallito@unimib.it

Leonor Josefina Romero Lauro
Department of Psychology & NeuroMi, University of Milano-Bicocca
leonor.romero1@unimib.it

Carlo Cecchetto
SFL (CNRS and Paris 8 University) and University of Milano-Bicocca
carlo.cecchetto@unimib.it

Received: 15-10-2022
Accepted: 18-01-2023
Published: 24-01-2023

How to cite: Sala, Marco, Bettelli, Giulia, Giustolisi, Beatrice, Vergallito, Alessandra, Romero Lauro, Leonor Josefina & Carlo Cecchetto. 2023. Comprehension of double-center embedded relatives in Italian: a case for hierarchical intervention. *Isogloss. Open Journal of Romance Linguistics* 9(1)/1, 1-29.

Abstract

Object relatives are more difficult to process than subject relatives. Several sentence processing models have been proposed to explain this difference. As double-center embedding relatives contain several long-distance dependencies, they are an ideal configuration to compare sentence processing models. The main aim of the present study was to compare the predictions of the featural Relativized Minimality approach with the ones of other relevant sentence processing models.

57 Italian-speaking healthy adults answered comprehension questions concerning the first, second, or third verb to appear in both double-center embedding and control sentences. Results show that questions concerning the matrix verb of double-center embedding structures were significantly easier and were associated with faster response times than questions concerning the embedded verbs. Furthermore, in object double-center embedding relatives the questions concerning the verb of the most embedded clause were easier than the ones concerning the verb of the intermediate embedded clause.

This pattern of results is consistent with featural Relativized Minimality but cannot be fully explained by other sentence processing models.

Keywords: sentence processing, object relatives, double-center embedding, Italian, hierarchical intervention.

1. Introduction

In psycholinguistic research, relative clauses have been a prosperous source of information about the cognitive operations involved in language processing. Subject and object relatives are usually distinguished according to the position of the gap[1]. In the subject relative in (1), the gap sits in the subject position of the relative clause. Instead, in the object relative in (2), it sits in the object position.

(1) The dog_1 [that e_1 is watching the cat] is running.
(2) The dog_1 [that the cat is watching e_1] is running.

Beyond the grammatical role of the antecedent, relative clauses can differ according to their position in the sentence. Relative clauses embedded between the subject and the verb of the main clause, such as those in (1) and (2), are center-embedded. However, relative clauses can also be placed peripherally to the right of the main clause, as in (3) and (4).

(3) The owner is searching for the dog_1 [that e_1 is watching the cat].
(4) The owner is searching for the dog_1 [that the cat is watching e_1].

In the literature, a consistent finding is that object relatives, particularly center-embedded, are more difficult to process than subject relatives (e.g., King & Just, 1991;

[1] In this paper we will use the term ‘gap’, which is more common in psycholinguistics to refer to the category also called ‘trace’ or ‘copy’ in formal syntax papers, and we will use the term ‘filler’ or ‘antecedent’ to refer to the category the gap depends on. The link between filler and gap is indicated by a subscript.

Gibson, 1998; Traxler, Morris, & Seely, 2002; Grodner & Gibson, 2005). Different types of processing models have been proposed to explain this result. These models can be divided into linear and hierarchical intervention models.

1.1. Linear intervention models

The intervention models based on working memory constraints explain the processing difficulty of object relatives by referring to the greater demands posed to working memory by these sentences. They typically refer to the linear order of words in the sentence, for example by measuring the distance between two elements in terms of the words that intervenes between these two elements.

In this context, the two leading theories of sentence complexity are Gibson's Dependency Locality Theory (Gibson, 2000) and Gordon's similarity-based approach (Gordon, Hendrick & Johnson, 2001; 2004).

1.1.1. Dependency Locality Theory

According to the Dependency Locality Theory, sentence comprehension is based on decay. This theory explains the processing difficulty by referring to two components of language comprehension that require working memory resources: a.) structural integration cost: incoming lexical items must be integrated into the syntactic representation of the sentence; b.) storage cost: the structure must be kept in memory and, in particular, it is necessary to keep track of incomplete dependencies. A fundamental principle of Gibson's approach is that the cost of integrating two elements or of maintaining a prediction about the syntactic structure depends on the linear distance between them. More precisely, distance is defined in terms of intervening discourse referents, where intervention is determined linearly. The activation of the stored representations decays when referential processing is required, making it more challenging to maintain predictions on the syntactic structure and integrate previously elaborated lexical items. Therefore, object relatives such as (5) are associated with more computational costs than subject relatives such as (6) since, in the former, a discourse referent ('the child') linearly intervenes between the antecedent and its gap. Instead, in (6), no discourse referents intervene between 'mother' and its gap.

(5) The mother$_1$ [that the child is calling e_1] is working.
(6) The mother$_1$ [that e_1 is calling the child] is working.

1.1.2. The similarity-based approach

Gordon, Hendrick, and Johnson (2001; 2004) proposed a theory of integration costs based instead on similarity. According to this model, the similarity of the intervening element with the filler is responsible for the greater processing difficulty of object relatives. Although intervention is not explicitly defined, one can infer that **B** is supposed to intervene between **A** and **C**, if **B** follows **A** and precedes **C**. Indeed, more processing resources are required to integrate the antecedent when an element qualified as a plausible candidate for the dependency intervenes. Gordon et al. (2001; 2004) offer some experimental support to this claim: for example, they show that a sentence like (7) in which the intervening element and the filler are both proper names is harder to process than a sentence like (8) where the intervening element is a pronoun.

(7) It was Bill$_1$ that Dan avoided e_1 at the party.

(8) It was $Bill_1$ that you avoided e_1 at the party.

Gordon et al. do not provide a precise theory of similarity since they do not specify which dimensions of similarity should be considered.

1.2. Hierarchical intervention models

The models considered so far assume, implicitly or explicitly, a linear (as opposed to hierarchical) definition of intervention. Since this point is crucial in our work, we define what we mean by hierarchical intervention. Following standard practice (cf. Rizzi, 1990 and much following work), hierarchical intervention is defined in terms of c-command, as in (9).

(9) Z intervenes between X and Y if Z c-commands Y and Z does not c-command X.

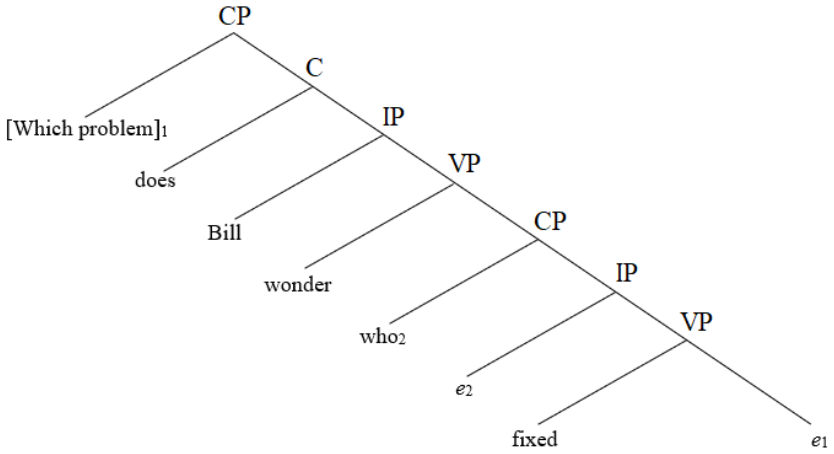
In turn, the c-command can be defined as follows (cf. Reinhart, 1976): A c-commands B if A does not dominate B and every node that dominates A also dominates B. Dominance is defined as a node A dominates a node B if one can trace a path from A to B by moving only downwards in the syntactic tree representing a sentence. As a consequence of this definition, a node c-commands its sister node and all of its sister's descendants in the syntactic tree.

The pattern in (10) and (11) can illustrate the difference between the linear and the hierarchical notions of intervention. From a baseline sentence like (10a), extraction of the 'which problem' causes degradation, as in (10b). In (10b) 'who' intervenes between 'which problem' and its gap both linearly ('who' is pronounced after 'which problem' but before the position in which a direct object is normally found, cf. (6)) and hierarchically ('who' c-commands the position of the gap but does not c-command the position of the filler, cf. the tree in (10c)).[2]

(10) a. Bill wonders who fixed the problem.
b. *$[\text{Which problem}]_1$ does Bill wonder who_2 e_2 fixed e_1?

[2] The syntactic trees in this paper are simplified in some respects as their main purpose is to visually show whether a potential intervener c-commands the gap in a filler-gap dependency or not. Following standard practice, when a phrase is not fully decomposed it is bracketed. We assume that *wh*-phrases move to Spec,CP and the auxiliary sits in C in interrogatives (for our purposes in this paper it is not necessary to delve into the details of left periphery as investigated by Rizzi, 1997). Another simplification is that the middle field (where Tense, Agreement and Mood is expressed) is not represented in the syntactic tree. Although our main point does not hinge on these specific implementations, for concreteness we represent the Chinese relativer *de* as a complementizer and we assume a raising analysis for relatives, in which the noun relabels the structure it merges with (see Cecchetto and Donati 2015 for motivation for this approach to relativization).

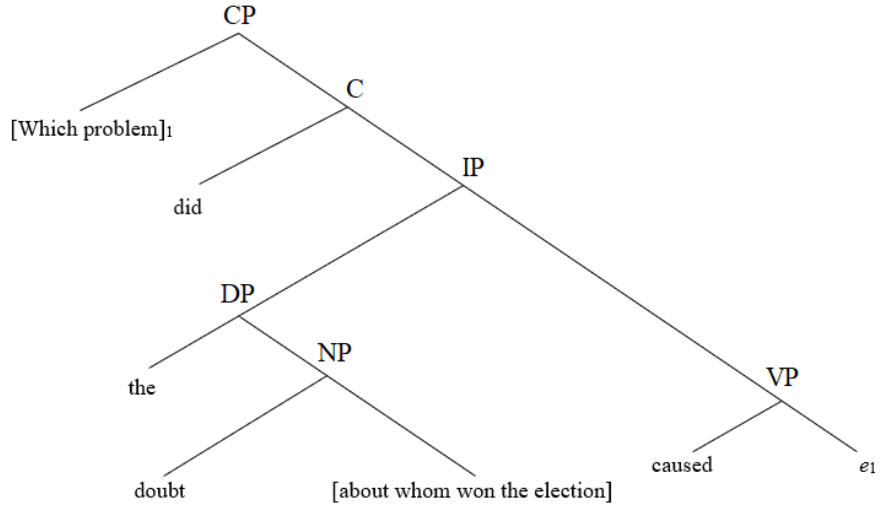
c.



As 'which problem' and 'who' share a relevant morphosyntactic feature (they are both *wh*-expressions), 'who' qualifies as an intervener.
(11) shows that the notion of intervention that counts is hierarchical rather than linear. From a baseline like (11a), movement of 'which problem' is possible, as in (11b). The reason is that 'who' intervenes only linearly, not hierarchically, as it is buried within the subject phrase, as shown in the syntactic tree in (11c). Under the definition of intervention in (9), 'who' is not an intervener in (11b), hence the structure is well-formed.

(11) a. The doubt about who won the election caused a big problem.
b. [Which problem]$_1$ did [the doubt about whom won the election] cause e_1?

c.



1.2.1. Active Filler Strategy

One of the first hierarchical intervention models has been proposed by Frazier and Clifton (1989), and Frazier and Flores D'Arcais (1989), which explained the greater

processing difficulty of object relatives by referring to parsing strategies, such as the Active Filler Strategy (Frazier, 1987). This model assumes that processing difficulties depend on the presence of gaps in the hierarchical representation. The Active Filler Strategy states that moved constituents are assigned to the first available gap. Therefore, in the presence of a relative pronoun, the antecedent should be treated as the subject of the relative clause. This parse is compatible with subject relatives but not with object relatives. In the latter case, a reanalysis will be necessary to assign the role of the object to the antecedent.

1.2.2. The Relativized Minimality based model

The sentence in (10b) is a typical example used to illustrate Rizzi's (1990) Relativized Minimality model, which is based on the hierarchical notion of intervention. An extension to psycholinguistics has been proposed by Friedmann, Belletti, and Rizzi (2009), who use the label "featural Relativized Minimality". This approach was first developed to explain why children can process subject - but not object - relatives until a certain age. While in classical Relativized Minimality, the feature that triggers intervention is a *wh*-feature (cf. (10)), under featural Relativized Minimality, it is assumed that a more articulate set of morphosyntactic features can cause interference. For example, in a sentence that contains an object relative such as (12), there are two relevant features to consider: the -/+R(elative) feature, a morphosyntactic feature that may manifest as the *wh*-traits and triggers the displacement of the object to the left peripheral position, and the -/+NP feature, which indicates the presence of a lexical restriction. Correspondingly, (12) is not processable by younger children because an intervener ('the owner') shares the NP feature with the filler. On the other hand, in the object relative in (13), the antecedent does not share the NP feature with the embedded subject (since the embedded subject is a pronoun with no lexical restriction). In this regard, Friedmann, Belletti, and Rizzi (2009) observed that when the antecedent and the embedded subject are featurally dissimilar (as in (13)), the comprehension of object relative clauses, which was pretty poor in (12), improved significantly in 22 Hebrew-speaking children aged between 3.7-5 years.

(12) The dog_1 (+R; +NP) that the owner (+NP) is watching e_1 is eating.
(13) The dog_1 (+R; +NP) that you (*pro*) are watching e_1 is eating.

Friedmann, Belletti, and Rizzi (2009) extended this explanation to the well-known processing difficulty of object relatives in adults, suggesting that adults would adhere to a less restrictive version of Relativized Minimality than young children. Indeed, the working memory resources necessary to process this syntactic structure are not yet developed in children. Therefore, children would need a disjunction between the morphosyntactic features of the antecedent and those of the intervening element to understand these sentences correctly. Conversely, adults have sufficient resources to process the sentence, but a partial overlap of morphosyntactic features, like in (12), would introduce a processing effort.

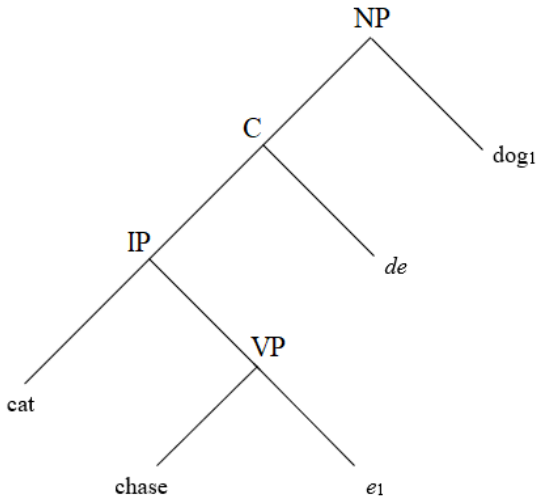
It is difficult to decide whether sentence processing relies on linear or hierarchical intervention models using externally headed postnominal relatives like English ones (see (14)). Indeed, the subject of the relative clause 'the cat' intervenes both hierarchically and linearly between filler and gap.

(14) The dog_1 [that the cat is chasing e_1].

However, there are configurations where linear and hierarchical intervention accounts make opposite predictions. A case that has been explored in the psycholinguistic literature is externally headed prenominal relatives in languages like Mandarin Chinese (cf. Huang & Li, 2009: chapter 6 for a presentation of the properties of this structure). An example is schematically illustrated in (15), where 'de' indicates the modification marker occurring between the relative clause and the head noun in Mandarin Chinese. As (15a) shows, the subject does not linearly intervene between the antecedent and its gap. It intervenes hierarchically, though, as indicated in (15b).

(15) a. [cat chase e_1] *de* dog_1
'the dog that the cat is chasing'

b.



Under linear intervention models, the expectation is that object relatives should not be more complex than subject relatives in the configuration abstractly illustrated in (15b). However, accumulating empirical evidence (cf. Jäger et al., 2015 for Chinese; Yun et al., 2010 for Korean; and Miyamoto et al., 2013 for Japanese) point toward a subject relatives' advantage even in this configuration, and these findings are more consistent with hierarchical intervention accounts (cf. Hsiao & Gibson, 2003 and Gibson & Wu, 2013 for the opposite view and Vasishth et al., 2013 for a response).
All in all, featural Relativized Minimality exploits a similar intuition as Gordon's similarity-based approach, but in addition it tries to explain why among different dimensions of similarity (e.g., semantic, phonological, morphosyntactic, etc.), only certain morphosyntactic features are relevant to intervention effects.

1.3. Surprisal Theory

Finally, a theory of sentence processing that has received particular attention is Levy's (2008) Surprisal Theory. Unlike the models previously considered, this is an experience-based model that remains neutral on the linear-vs-hierarchical intervention debate. According to Surprisal Theory, expectations based on experience are formed at different linguistic levels during sentence processing. The cognitive load posed by

a linguistic element is a function of its probability given the previous context. Therefore, the greater the surprisal value of a linguistic element, the greater its processing difficulty (Hale, 2001; 2003; Levy, 2008). Object relatives are more difficult to process in this context than subject relatives because they are less frequent. Despite its popularity, the Surprisal Theory raises some questions. First, there is the issue of choosing the corpora upon which syntactic probability is computed (cf. Hale, 2001). Second, one can ask why object relatives are less frequent. Obviously, responding that they are less frequent because they are more difficult would be circular. Finally, Staub et al. (2018) offered experimental evidence that object relatives remain challenging even when they compete with an alternative parse (i.e., nominal complement clauses) that is less frequent, casting doubts on the role of frequency as an exhaustive explanatory factor for syntactic constructions.

1.4. Double-center embedded relatives

This paper investigates a specific case in which different approaches to sentence processing make different predictions, namely double-center embedded relatives. Double-center embedded relatives are characterized by the fact that, in addition to the relative clause embedded between the subject and the verb of the main clause, another relative clause modifies a nominal element belonging to the subordinate clause of a higher degree. These sentences can also be distinguished based on the grammatical role of the filler: in the subject double-center embedded relatives like (16), there are two subject relatives, while in the object double-center embedded relatives like (17), there are two object relatives.

(16) The boy$_1$ [that e_1 calls the dog$_2$ [that e_2 chases the cat]] eats ice cream.
(17) The boy$_1$ [that the woman$_2$ [that the dog watches e_2] scolds e_1] eats ice cream.

The processing of a double level of center embedding is notoriously very challenging. In this regard, several authors have observed that this is particularly true for the object double-center embedded relatives: in most cases, they cannot be processed online, and their meaning requires metalinguistic reasoning (Chomsky & Miller, 1963; Bever, 1970; Gibson, 1998; Karlsson, 2007).

In the present work, we investigated double-center embedding in Italian. We chose Italian because it is a language with externally headed relative clauses in which it is possible to build double-center embedding configurations. Our interest in these structures is that they contain different kinds of dependencies (filler-gap dependencies and main subject-verb) and multiple filler-gap dependencies. Indeed, the approaches to sentence complexity considered so far make different predictions regarding their processing.

To experimentally explore this issue, we asked participants to answer questions concerning the first, second, or third verb to appear in the sentence after hearing the relevant sentence (for a sentence like (16), the three questions would be respectively: 'who is called?', 'who is chased?', and 'who eats an ice cream?'). Experimental stimuli included sentences with a double level of center embedding, both subject (16) and object (17) relatives, and control sentences of comparable length but involving coordination rather than subordination (e.g., 'The boy eats an ice cream, the women a sandwich, the dog runs.' or 'The boy is eating while the women jump and the dog runs.').

The experimental material was developed to build a set of complex stimuli, in principle useful for answering further different empirical questions, different from those of this study.

1.5. Predictions of the different approaches to sentence processing.

We focus here on object double-center embedded relatives because they are an ideal configuration to compare the different hypotheses to explain the complexity of object relatives in general. In particular, by taking (17) as a representative sentence, the predictions of the different approaches can be seen in Table 1.

Table 1. Predictions for the different levels of difficulty that questions concerning object double-center embedded relatives should posit, according to different approaches to sentence complexity. More dots mean more difficulty.

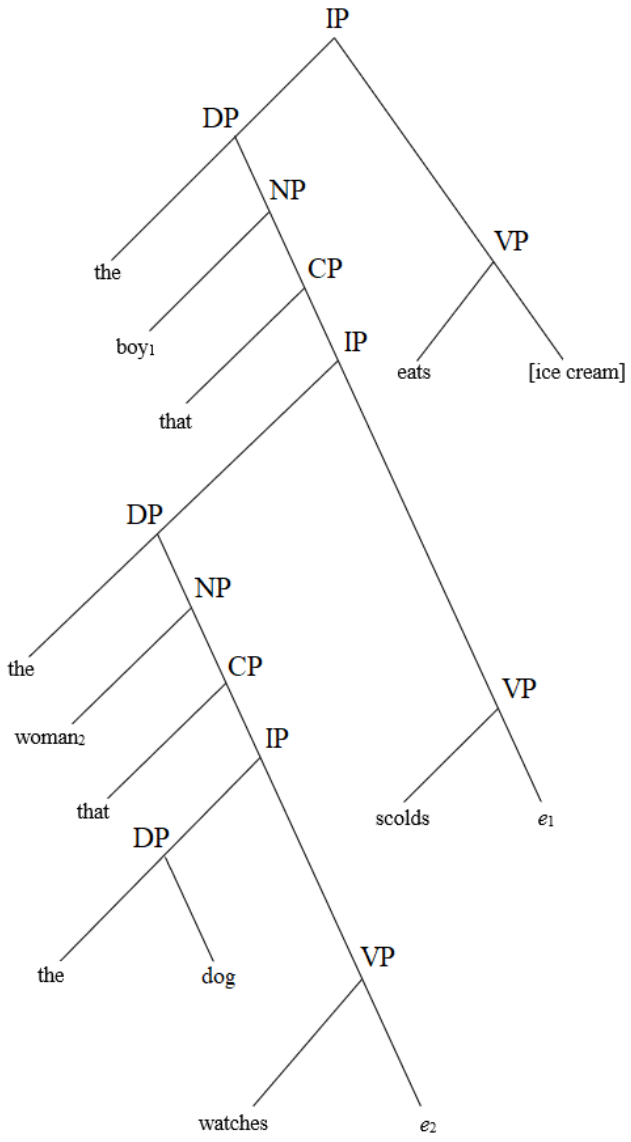
	Q1: Who is watched?	**Q2**: Who is scolded?	**Q3**: Who is eating?
Linear intervention approaches:			
Dependency Locality Theory	●●	●●●	●●●
Hierarchical intervention approaches:			
Active Filler Strategy	●●●	●●●	●
featural Relativized Minimality	●●	●●●	●
Neutral approaches:			
Surprisal Theory	●●●	●●	●

Dependency Locality Theory predicts that the easier question should be Q1. This is so because the filler-gap dependency concerning the first verb has the lowest linear distance between syntactic dependents (e.g., in (17) 'the dog' is the only NP that intervenes between 'the woman' and its gap). Instead, answer Q2 and Q3 require the processing of dependencies in which there is a greater linear distance between dependents (e.g., in (17) 'the woman' and 'the dog' intervene when the answer requires retrieving the NP 'the boy').

The Active Filler Strategy, which makes the presence of gaps a core source of processing difficulty, predicts that Q3 should be the most straightforward question since it requires the processing of the main subject-verb dependency but does not imply the elaboration of a gap. Instead, Q1 and Q2 should be more difficult since answering these questions requires processing a gap whose position in the sentence needs to be reanalyzed.

Also, featural Relativized Minimality predicts that Q3 should be the easiest question to answer. Indeed, although linearly the main subject and verb are very distant and two DPs intervene between them (i.e., 'the woman' and 'the dog'), hierarchically, there is zero distance between them, as shown in (18).

(18)



Featural Relativized Minimality also predicts that answering Q2 should be more complicated than answering Q1. Indeed, in the dependency concerning the second verb, there are more intervening elements with relevant morphosyntactic features such as lexical restriction (i.e., 'the woman' and 'the dog') compared to the dependency concerning the first verb (i.e., only 'the dog').

Although Levy (2008) does not explicitly discuss the case of long-distance subject-predicate dependencies, presumably, Surprisal Theory predicts that questions concerning the main subject-verb dependency, namely Q3, should be the easiest to answer since the main verb is expected as soon as the main subject is met. Furthermore, Surprisal Theory predicts that the answer to the question concerning the first verb, namely Q1, should be more difficult than the answer to the question concerning the second verb, namely Q2. Indeed, answering Q1 requires processing the second level

of embedding, which is more unexpected than the first level since double-center embedding is exceedingly rare (Karlsson, 2007).

Since object double-center embedded relatives are an ideal configuration to study this issue, this paper aims to compare the predictions of different accounts of sentence complexity to achieve a greater understanding of the factors involved.

2. Methods

2.1. Participants

60 (40 online, 20 in the lab) healthy, right-handed native Italian speakers took part in a two-sessions study (24 males, mean age = 23.5, SD = 2.2). The participants were resident in Italy, while the degree of bilingualism was not checked. Participants were recruited through the Sona System platform of the University of Milan-Bicocca. Therefore, the sample was mainly composed of university students (minimum level of education = secondary school diploma) who did not know the purpose of the study. The local ethical committee approved the study. Participants were treated following the ethical principles stated in the Declaration of Helsinki.

2.2. Materials

All materials are accessible through the Materials folder in the OSF repository of the current project (https://osf.io/7zg6m/).

The experimental set consisted of 2 types of target sentences, namely object double-center embedded relatives (Rel_Obj) and subject double-center embedded relatives (Rel_Subj), and 2 types of control sentences, namely control trials without a conjunction (Control) and control trials with 'and' and 'while' conjunctions (Control_While). Each type of sentence included 48 trials, for a total number of 192 sentences. Sentences were created from 12 nouns triplets, shared across target and control sentences, and 3 sets of 12 verb triplets (one for target sentences and two for control sentences, as detailed in the Materials table on the OSF repository). All sentences were composed of 12 words.

The four sentence types were the following:

1. Rel_Obj such as (19):

(19) Il giornalaio$_1$ [che le poliziotte$_2$ [che il meccanico critica e_2] coprono e_1] sta piangendo.
The newsagent$_1$ [that the policewomen$_2$ [that the mechanic critics e_2] cover e_1] is crying.

These are sentences with a double level of embedding interrupting the main clause's processing. The antecedent of the relative clause of a higher degree (to which we will refer to as the intermediate embedded clause, as it is in-between the main clause and the most embedded clause) '(il) giornalaio', is the subject of the main sentence ('Il giornalaio sta piangendo' – the newsagent is crying) and corresponds to the object gap inside the same relative clause ('Le poliziotte coprono il giornalaio' – the policewomen cover the newsagent). The antecedent of the most embedded relative clause, '(le) poliziotte', also corresponds to the object gap ('Il meccanico critica le poliziotte' – the mechanic criticizes the policewomen).

2. Rel_Subj such as (20):

(20) La signora$_1$ [che e_1 indica i ragazzi$_2$ [che e_2 inseguono la bambina]] sta tremando.
The woman$_1$ [that e_1 indicates the boys$_2$ [that e_2 chase the little girl]] is shaking.

Even in this configuration, a double level of embedding interrupts the processing of the main clause. However, here the antecedent of the relative clause of a higher degree (called also in this case intermediate embedded clause, for the same reason explained in the previous paragraph) '(la) signora', is the subject of the main clause ('La signora sta tremando' – the woman is shaking) that corresponds to the subject gap ('La signora indica i ragazzi' – the woman indicates the boys). As for the antecedent of the most embedded relative clause '(i) ragazzi', it also corresponds to the subject gap ('I ragazzi inseguono la bambina' – the boys chase the little girl).

3. Control such as (21):

(21) Il nonno ascolta musica classica, i signori __ musica rock, la ragazza lavora.
The grandfather listens to classical music, the men __ to rock music, the girl works.

These are sentences composed of three coordinated clauses without explicit conjunction. The subject of the second clause performed the same action (expressed by a transitive verb) as the subject of the first clause, although to a different object. In the second clause, the verb is omitted (namely, this is a case of gapping in the sense of Ross, 1967). Lastly, the subject of the third clause performed a different action (expressed by an intransitive verb) compared to the subjects of the first and second clauses.

4. Control_While such as (22):

(22) Il meccanico sta cavalcando mentre le poliziotte mormorano e il giornalaio gioca.
The mechanic is riding while the policewomen whisper and the newsagent plays.

These are sentences composed of three coordinated clauses. The coordination between the first and second clauses has been realized through the conjunction 'while' (in Italian: 'mentre'). In contrast, the coordination between the second and third clauses has been realized through the conjunction 'and' (in Italian: 'e'). These coordinated clauses were composed of independent subjects carrying out different actions (expressed by intransitive verbs).

Each type of sentence included clauses with noun phrases following a singular-plural-singular pattern, such as those illustrated in the previous examples (19), (20), (21), and (22), and sentences with noun phrases following a plural-singular-plural pattern, such as (23), (24), (25), and (26), in equal number. The number mismatch was aimed at reducing the potential ambiguity of the sentences.

1. Rel_Obj:
(23) I giornalai che la poliziotta che i meccanici criticano copre stanno piangendo.
The newsagents that the policewoman that the mechanics criticize covers are crying.

2. Rel_Subj:
(24) Le signore che indicano il ragazzo che insegue le bambine stanno tremando.
The women that indicate the boy that chases the little girls are shaking.

3. Control:
(25) I nonni ascoltano musica classica, il signore musica rock, le ragazze lavorano.
The grandfathers listen to classical music, the man to rock music, the girls work.

4. Control_While:
(26) I meccanici stanno cavalcando mentre la poliziotta mormora e i giornalai giocano.
The mechanics are riding while the policewoman whispers and the newsagents play.

All sentences, pronounced by an adult Italian female, were audio recorded in a soundproofed lab and edited by a professional technician whose aim was to make the audio as clear as possible (all audio files are collected in the Materials folder in the OSF repository of the current project). Prosody was checked so that sentences were pronounced naturally.
Sentence comprehension was assessed through a written question (see Table 2). Each sentence was matched with three different questions, each concerning one of the three verbs of the sentence. Each participant received one of the three possible questions for each sentence.

Table 2. Example of the questions used to assess comprehension for each sentence type.

Sentence type	**Sentence**	**Question**		
		1	**2**	**3**
Rel_Obj	Il giornalaio [che le poliziotte [che il meccanico critica$_1$] coprono$_2$] sta piangendo$_3$. – The newsagent [that the policewomen [that the mechanic critics$_1$] cover$_2$] is crying$_3$.	Chi viene criticato? – Who is criticized?	Chi viene coperto? – Who is covered?	Chi sta piangendo ? – Who is crying?
Rel_Subj	La signora [che indica$_1$ i ragazzi [che inseguono$_2$ la bambina]] sta tremando$_3$. – The woman [that indicates$_1$ the boys [that chase$_2$ the little girl]] is shaking$_3$.	Chi viene indicato? – Who is indicated?	Chi viene inseguito? – Who is chased?	Chi sta tremando? – Who is shaking?
Control	Il nonno <u>ascolta musica classica</u>$_1$, i signori <u>musica</u>	Da chi è ascoltata	Da chi è ascoltata	Chi lavora?

	rock$_2$, la ragazza lavora$_3$. – The grandfather listens to classical music$_1$, the men to rock music$_2$, the girl works$_3$.	musica classica? – [*lit*. By whom is classical music listened to?]	musica rock? – [*lit.* By whom is rock music listened to?]	– Who works?
Control_ While	Il meccanico sta cavalcando$_1$ mentre le poliziotte mormorano$_2$ e il giornalaio gioca$_3$. – The mechanic is riding$_1$ while the policewomen whisper$_2$ and the newsagent plays$_3$.	Chi sta cavalcando? – Who is riding?	Chi mormora? – Who whispers?	Chi gioca? – Who plays?

The three questions (1, 2, 3) were ordered according to the linear position in the sentence of the verb presented in the question. In particular, question 1 concerns the first verb appearing in the sentence (marked by the subscript 1 in each sentence in Table 2), question 2 concerns the second verb (marked by the subscript 2), and question 3 concerns the third verb (marked by the subscript 3).

Notice that in Rel_Obj sentences, question 1 concerns the verb of the most embedded clause, question 2 the verb of the intermediate embedded clause, and question 3 the verb of the main sentence. In Rel_Subj sentences, question 1 concerns the verb of the intermediate embedded clause, question 2 the verb of the most embedded clause, and question 3 the verb of the main sentence.

The two questions concerning the embedded verbs used a passive form to avoid ambiguity for the double-center embedded relative clauses. For this reason, we used a passive form also to assess the comprehension of Control sentences (question 1 and question 2, concerning a transitive verb).

Considering the two types of target sentences, questions were always composed of three words, the interrogative pronoun 'who' (in Italian: 'chi') and the verbs. Considering the two types of control sentences, questions' length differed from two to six words.

2.3. Procedure

The task was an auditory comprehension task. After listening to a sentence, the participant had to answer a written question that appeared at the top of the screen together with three answer options presented below (see Figure 1). Participants were asked to choose the correct answer by pressing a key on the keyboard ("1" for the first option, "2" for the second, and "3" for the third one). They had a time limit of 10 seconds to pick the proper answer. After the participants' choice, the following item was presented.

Figure 1. Graphic representation of the experimental procedure.

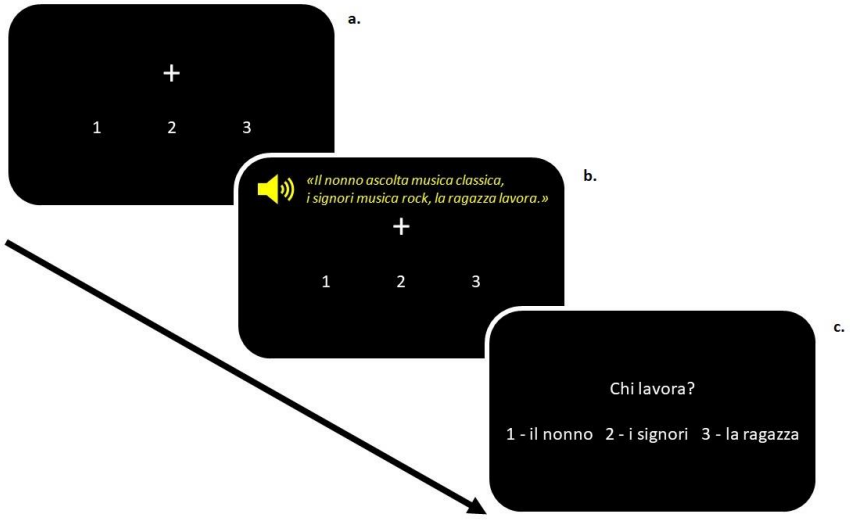


Figure 1. The experimental procedure consisted in: a. Fixation point presentation; b. Sentence auditory presentation (in this example, 'Il nonno ascolta musica classica, i signori musica rock, la ragazza lavora.' – the grandfather listens to classical music, the men to rock music, the girl works); c. presentation of the written question ('Chi lavora?' – who works?) together with the 3 answer options ('il nonno' – the grandfather; 'i signori' – the men; 'la ragazza' – the girl).

Before starting the experiment, written instructions were given to participants, with the recommendation to respond as accurately and quickly as possible, positioning the index, middle, and ring finger of their right (dominant) hand above the keys 1, 2, and 3, respectively.

All participants took part in two experimental sessions. At the beginning of each session, participants were presented with initial training to get familiar with the task. The training comprised 10 new sentences, 8 covering the four conditions included in the experimental dataset plus 2 sentences with coordination in which each subject was matched with a predictable action.

The two experimental sessions were held at a minimum distance of 2 days from each other and a maximum distance of 7 days (online participants received the link to the second session a few days after completing the first one). The experimental material was therefore divided into two lists to avoid a long session (A and B; one list per experimental session) containing 96 sentences each, 24 per type. In each list, sentences were divided into three blocks, each comprising 8 sentences per type. Blocks were separated by a break, whose length was decided by the participants, who could resume the experiment whenever they wanted. Lists and blocks presentation order was counterbalanced across participants, whereas items order was randomized within each block. Each session lasted about 25 minutes, and students received course credits for their participation. Accuracy and reaction times (RTs) were recorded.

Due to the SARS-COVID-19 pandemic, data collection occurred initially remotely and then in lab. Online participants (data collected between March and July 2021) performed the experiment using the E-prime Go platform built for E-Prime 3 (Psychology Software Tools Inc., Pittsburgh, PA, United States). E-prime Go allows sharing the experiment via a link, and participants can run the experiment on their pc.

Participants were asked to sit in a quiet room and wear headphones to ensure they could listen appropriately to the stimuli.
In-lab participants (data collected between October and November 2021) were tested in a quiet room and wore headphones. The experiment was run on E-Prime 3 software.

2.4. Statistical procedure
Statistical analyses were carried out using the statistical software R (R Core Team, 2021). The complete dataset and analyses scripts are in the OSF repository of the current project (https://osf.io/7zg6m/).
First, we performed a preliminary check of accuracy in both Control and Control_While sentences. Three participants showed a mean accuracy below 80% in the Control sentences (one of them also had an accuracy below 80% in the Control_While sentences, see Figure 2) and were removed from the subsequent analyses. Ultimately, we analyzed the results of 57 participants (24 males, mean age = 23.5, SD = 2.2).

Figure 2. Participants mean accuracy in the different sentence types.

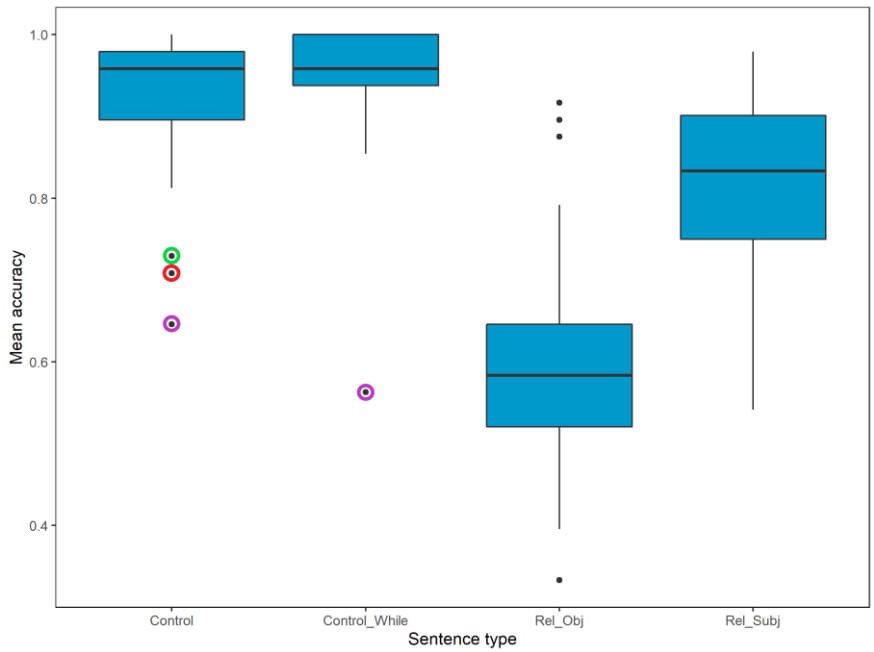


Figure 2. Outliers detected in the control sentences are circled. Two outliers (green and red circles) had a low performance in the Control sentences, while one outlier (violet circles) had a low performance in both Control and Control_While conditions.

Two mixed model regressions were run using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015). The first model had the dichotomous variable accuracy (0 = wrong; 1 = correct) as a dependent variable; therefore, the *glmer* function for generalized mixed models was used. The second model's dependent variable was RTs (measured in milliseconds). Only RTs associated with a correct answer were considered for analyses (83% of the trials). Outliers were removed via model criticism (2.5 *SD* of standardized residuals: Baayen, 2008; Ch. 6.2.3). The remaining data (81%) were analyzed with the *lmer* function for linear mixed models. All analyses were

conducted after excluding the responses provided by the participants to the training items.

Both models originally included the following fixed factors: *(i)* question (three levels: 1 vs. 2 vs. 3), *(ii)* sentence type (four levels: Control vs. Control_While vs. Rel_Obj vs. Rel_Subj), *(iii)* blocks order (three levels: first vs. second vs. third), *(iv)* lists order (two levels: A-B vs. B-A), and *(v)* experiment modality (two levels: online vs. in-presence). Moreover, the interactions between *(i)* question and *(ii)* sentence type and the one between *(iii)* blocks order and *(iv)* lists order were entered in the original model. The fixed factors *(iii)*, *(iv)*, and *(v)* have been considered to control for the possible differences in data collection. Indeed, we aimed at checking if the performances of the participants might be modulated by: the order of presentation of the blocks (i.e., if performances changed while progressing from the first to the third block of each list), the order of lists presentation (i.e., if performances were different depending on which list they were first exposed to), and the different modality (i.e., if there were differences between online and in-presence performances). Notably, the interaction between *(i)* and *(ii)* was the main focus of our analysis. In particular, we aimed to see how participants answered the different types of *(i)* questions across the different *(ii)* sentence types. As we previously highlighted (see 2.2), each question concerns the verb of a different clause in the different sentence types.

To identify the maximal converging model for each dependent variable, we used the *buildmer* function (*buildmer* package, Voeten, 2022). Participants and Items were included as random intercepts to account for participant-specific variability and item-specific idiosyncrasies (Baayen et al., 2008), as well as by-item random slopes for the effect of the different sentence types. All the post-hoc comparisons were run through the *testInteractions* function of the *phia* package (De Rosario-Martinez, 2015), and *p* values were corrected with the False Discovery Rate (FDR; Benjamini & Hochberg, 1995).

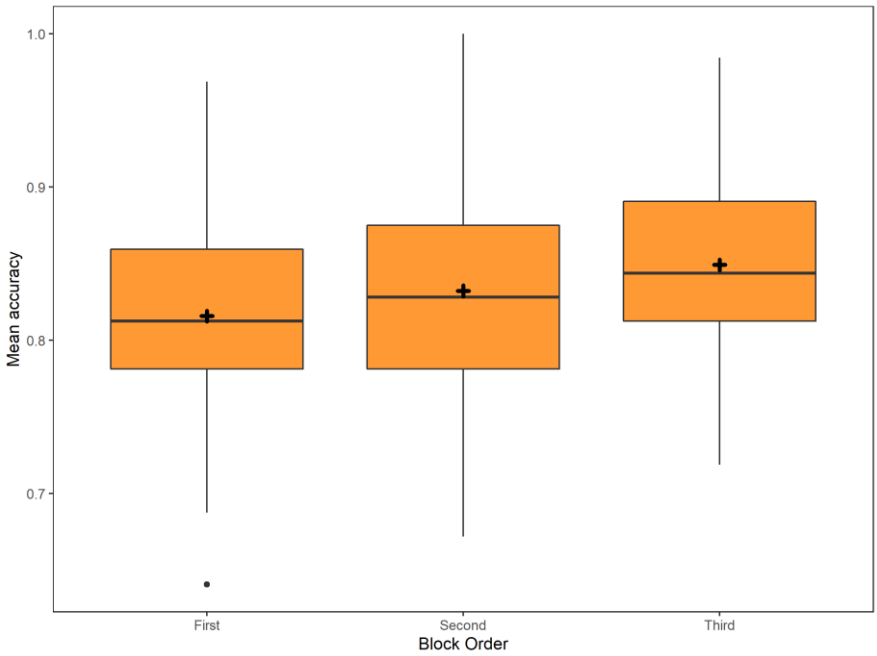
3. Results

3.1 Accuracy

The best fitting model for accuracy included the effects of the blocks order ($\chi2_{(2)}$ = 21.95, p <.001), question ($\chi2_{(2)}$ = 633.80, p <.001), sentence type ($\chi2_{(3)}$ = 907.27, p <.001), and the question by sentence type interaction ($\chi2_{(6)}$ = 161.23, p <.001).

Concerning the simple effect of blocks order, post-hocs showed that accuracy was significantly higher in the second block compared to the first (β = 0.458, df = 1, $\chi2$ = 5.033, p = .025) and in the third block compared to the first (β = 0.411, df = 1, $\chi2$ = 21.950, p <.001) and the second (β = 0.452, df = 1, $\chi2$ = 6.051, p = .021) (see Figure 3).

Figure 3. Mean accuracy of the participants in the different blocks.



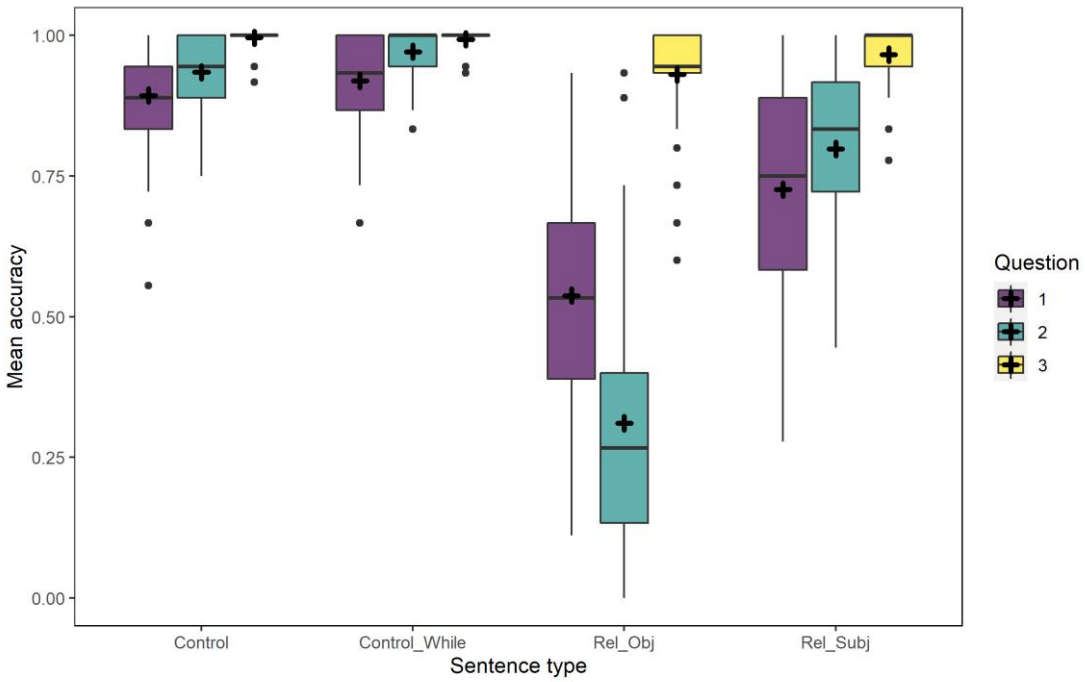
Analyzing the interaction between sentence type and question, we found that in Rel_Obj sentences, accuracy was significantly higher in question 3 (concerning the verb of the main clause) compared to question 1 (concerning the verb of the most embedded clause) ($p < .001$) and question 2 (regarding the verb of the intermediate embedded clause) ($p < .001$). In this sentence type, accuracy was also significantly higher in question 1 (verb of the most embedded clause) than in question 2 (verb of the intermediate embedded clause) ($p < .001$).

In Rel_Subj sentences, accuracy was significantly higher in question 3 (verb of the main clause) compared to question 1 (verb of the intermediate embedded clause) ($p < .001$) and question 2 (verb of the most embedded clause) ($p < .001$). Accuracy was also significantly higher in question 2 (verb of the most embedded clause) than in question 1 (verb of the intermediate embedded clause) ($p < .001$).

Lastly, in Control and Control_While sentences, accuracy was significantly higher in question 3, compared to question 1 (Control: $p < .001$; Control_While: $p < .001$) and question 2 (Control: $p < .001$; Control_While: $p < .001$), and in question 2 compared to question 1 (Control: $p < .001$; Control_While: $p = .001$).

See Table 1 in the Supplementary Materials for the post-hoc results and Figure 4.

Figure 4. Mean accuracy of the participants in each sentence type divided by question.



Control: Il nonno ascolta **musica classica**$_1$, i signori ___ **musica rock**$_2$, la ragazza **lavora**$_3$.
The grandfather listens to classical music, the men to rock music, the girl works.

Control_While: Il meccanico **sta cavalcando**$_1$ mentre le poliziotte **mormorano**$_2$ e il giornalaio **gioca**$_3$.
The mechanic is riding while the policewomen whisper and the newsagent plays.

Rel_Obj: Il giornalaio [che le poliziotte [che il meccanico **critica**$_1$] **coprono**$_2$] **sta piangendo**$_3$.
The newsagent that the policewomen that the mechanic critics cover is crying.

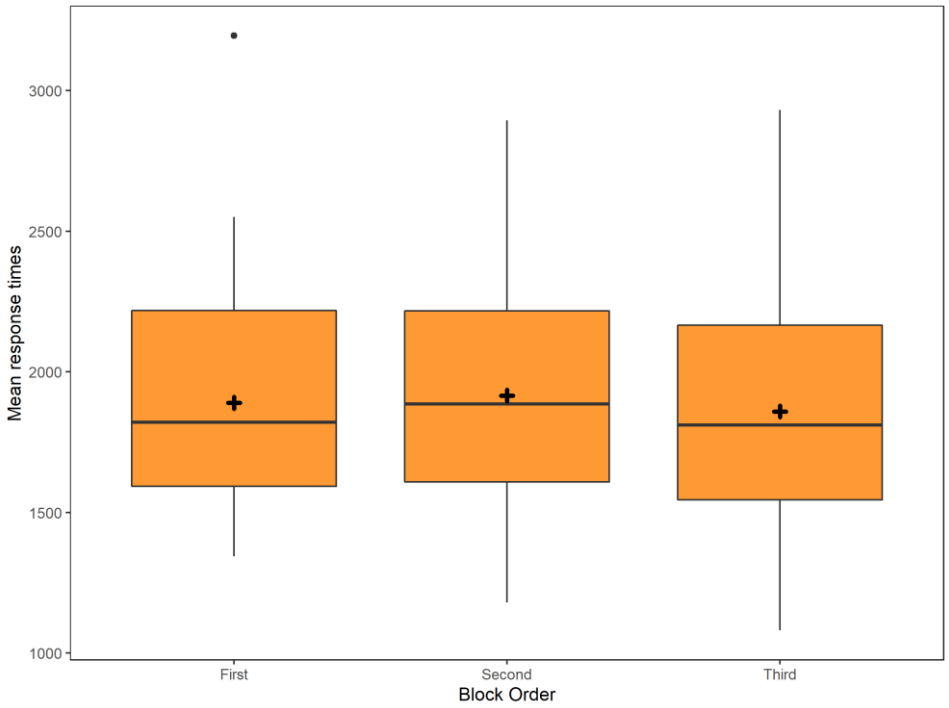
Rel_Subj: La signora [che **indica**$_1$ i ragazzi [che **inseguono**$_2$ la bambina]] **sta tremando**$_3$.
The woman that indicates the boys that chase the little girl is shaking.

3.2 Response times

The best fitting model for response times included the effects of the blocks order ($\chi2_{(2)}$ = 10.412, p = .005), question ($\chi2_{(2)}$ = 3586.526, p <.001), sentence type ($\chi2_{(3)}$ = 1167.709, p <.001), and the question by sentence type interaction ($\chi2_{(6)}$ = 515.100, p <.001).

Concerning the simple effect of blocks order, post-hocs showed that response times were significantly lower in the third block compared to the second (β = 65.840, df = 1, $\chi2$ = 10.381, p = .004). The other two comparisons (first vs. second and first vs. third) were not significant (see Figure 5).

Figure 5. Mean accuracy of the response times in the different blocks.



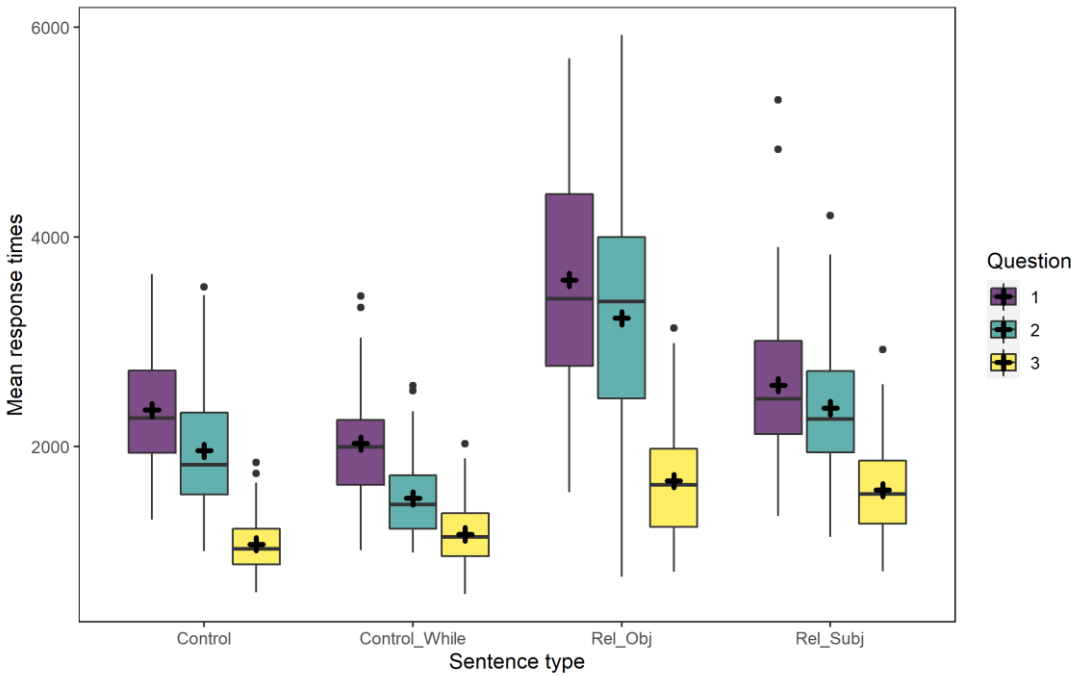
Analyzing the interaction between sentence type and question, we found that, for Rel_Obj sentences, response times were significantly faster in question 3 (verb of the main sentence) compared to question 1 (verb of the most embedded clause) ($p < .001$) and question 2 (verb of the intermediate embedded clause) ($p < .001$). In this type of sentence, response times were significantly faster in question 2 than in question 1 ($p = .001$).

As for Rel_Subj sentences, response times were significantly faster in question 3 (verb of the main sentence) compared to question 1 (verb of the intermediate embedded clause) ($p < .001$) and question 2 (verb of the most embedded clause) ($p < .001$). Moreover, response times were significantly faster in question 2 than in question 1 ($p = .001$).

In the two types of control sentences (Control and Control_While), RTs were significantly slower in answering question 1, compared to question 2 (Control: $p < .001$; Control_While: $p < .001$) and question 3 (Control: $p < .001$; Control_While: $p < .001$), and in question 2 compared to question 3 (Control: $p < .001$; Control_While: $p < .001$).

See Table 2 in the Supplementary materials for the post-hoc results and Figure 6.

Figure 6. Mean response times of the participants in each sentence type divided by question.



Control: Il nonno ascolta **musica classica**$_1$, i signori ___ **musica rock**$_2$, la ragazza **lavora**$_3$.
The grandfather listens to classical music, the men to rock music, the girl works.

Control_While: Il meccanico **sta cavalcando**$_1$ mentre le poliziotte **mormorano**$_2$ e il giornalaio **gioca**$_3$.
The mechanic is riding while the policewomen whisper and the newsagent plays.

Rel_Obj: Il giornalaio [che le poliziotte [che il meccanico **critica**$_1$] **coprono**$_2$] **sta piangendo**$_3$.
The newsagent that the policewomen that the mechanic critics cover is crying.

Rel_Subj: La signora [che **indica**$_1$ i ragazzi [che **inseguono**$_2$ la bambina]] **sta tremando**$_3$.
The woman that indicates the boys that chase the little girl is shaking.

4. Discussion

This paper aimed to investigate factors playing a crucial role in the processing difficulty observed with object relatives. We focused on object double-center embedded relatives because different accounts of sentence complexity make opposite predictions in this configuration.

We recruited 60 Italian-speaking healthy adults who listened, in two separate experimental sessions, to 192 acoustically presented pre-recorded sentences. After each sentence, a written question followed by three possible response options appeared. Questions were related to the three possible verbs of 4 different sentence types: object and subject double-center embedded relatives (named Rel_Obj and Rel_Subj, respectively), and 2 control sentences (involving two kinds of coordination, with and without a conjunction, called Control_While and Control, respectively).

Accuracy and RTs were recorded and separately analyzed through mixed model regressions. Concerning accuracy, the best fitting model included the simple effect of the blocks' order and the interaction between sentence types and questions. The order of the blocks showed that participants' accuracy increased while performing the task, thus suggesting a learning effect. Considering the interaction between sentence types and questions, the results showed that for Rel_Obj and Rel_Subj,

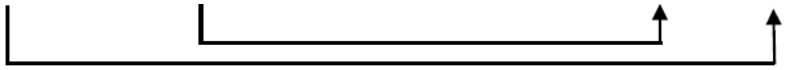
questions concerning the verb of the intermediate embedded clause (respectively, question 2 for Rel_Obj and question 1 for Rel_Subj) had the lowest accuracy, while questions concerning the verb of the most embedded clause (respectively, question 1 for Rel_Obj and question 2 for Rel_Subj) were those of intermediate difficulty. The questions concerning the last verb to appear (question 3 for both types of sentences) had the highest accuracy. Concerning the control sentences (Control and Control_While), accuracy was higher for the questions targeting the last verb to appear. In contrast, questions targeting the first and the second verb to appear were, respectively, the most difficult and the intermediate ones.

Concerning RTs, the best fitting model was the one including the interaction between sentence types and questions. For Rel_Subj, all results in RTs reflect the ones shown for accuracy, with slower RTs for questions concerning the verb of the intermediate embedded clause (question 1) compared to the verb of the most embedded (question 2) and the verb of the main clause (question 3), and for questions concerning the verb of the most embedded compared to the verb of the main clause. For Rel_Obj, in line with the results on the accuracy, RTs were slower for questions concerning the verb of the intermediate embedded (question 2) and the most embedded (question 1) clause compared to the verb of the main clause (question 3). However, RTs were slower for questions on the most embedded verbs than the intermediate embedded clause, unlike what we found for accuracy. This discrepancy will be addressed later in this Discussion. For the control sentences, differences between the conditions reflect the results found for accuracy. Indeed, RTs were faster for questions concerning the last verb than questions related to the first and second verbs and for the second compared to the first verb.

Interestingly, the best fitting models for the data did not include experiment modality (online vs. in-presence) as a fixed factor. Therefore, the experimental stimuli set allowed us to collect data online as reliably as in the lab.

The most relevant results for the topic addressed in this paper are those concerning Rel_Obj sentences, such as (27). As seen in paragraph 1.5, different accounts of sentence complexity make different predictions regarding processing the main subject-verb dependency in these sentences. Indeed, linear accounts of intervention such as Gibson's Dependency Locality Theory predict that this dependency should be complicated to process since the main subject and verb are linearly very far. Instead, hierarchical accounts of intervention such as featural Relativized Minimality predict that this dependency should be the easiest to process since hierarchically there is zero distance between the main subject and verb (Active Filler makes a similar prediction as this dependency does not involve a gap). Even Surprisal Theory, which remains neutral on the issue of linear as opposed to hierarchal intervention, makes this prediction since the main verb is expected as soon as the main subject is met.

(27) Rel_Obj:
The newsagent$_1$ that the policewomen$_2$ that the mechanic critics e_2 covers e_1 is crying.



In our experiment, the processing of the main subject-verb dependency was investigated with questions concerning the third verb ('who is crying? The

newsagent'), namely 'Q3'. As shown above, the questions concerning the third verb were the easiest to answer.

The same predictions regarding the main subject-verb dependency can be applied to Rel_Subj sentences such as (28). Again, the questions concerning the third verb ('who is shaking? The woman') were the easiest to answer.

(28) Rel_Subj:
The woman$_1$ that e_1 indicates the boys$_2$ that e_2 chase the little girl is shaking.

Different accounts of sentence complexity make different predictions also regarding the processing of the embedded clauses in Rel_Obj. As seen in paragraph 1.5, linear intervention models such as Gibson's Dependency Locality Theory predict that the intermediate embedded clause should be more difficult to process than the most embedded clause since, in the former, there is a greater linear distance between syntactic dependents. Even featural Relativized Minimality makes this prediction since the intermediate embedded clause requires the processing of a dependency in which there are more intervening elements with relevant morphosyntactic features ('the policewomen' and 'the mechanic' in (27)) compared to the most embedded clause (only 'the mechanic' intervenes in (27)). The Active Filler Strategy, which makes the presence of gaps a core source of processing difficulty, predicts no difference between these clauses since both require processing a gap whose position in the sentence needs to be reanalyzed. Finally, Surprisal Theory predicts that the most embedded clause should be more difficult to process than the intermediate embedded clause since the second level of embedding is rarer, therefore more unexpected, than the first level of embedding (whatever corpus one considers).

In our experiment, the processing of the most embedded and the intermediate dependency was investigated with questions concerning respectively the first ('who is being criticized? The policewomen') and the second verb ('who is being covered? The newsagent'), namely Q1 and Q2. As shown above, Q2 is answered less accurately than Q1. However, this disadvantage for Q2 was not reflected in response times, which were faster with Q1. We explain this discrepancy considering that participants answered randomly to questions related to the intermediate embedded clause in this type of sentences. Consequently, faster reaction times in this condition should reflect the fact that participants put less effort into retrieving the correct answer, not because this was easier but because it was too complicated, and they gave up (as discussed in the Introduction, the double level of center embedding is known to be extremely hard to be processed online).

Overall, our results support a notion of hierarchical intervention, specifically featural Relativized Minimality. The advantage of featural Relativized Minimality over accounts of intervention such as Gibson's Dependency Locality Theory is that the latter do not explain the ease with which the main subject-verb dependency is processed. This dependency should be complicated to process since in sentences with double-center embedding, the main subject and verb are linearly very far. The advantage of featural Relativized Minimality over accounts as the Active Filler Strategy and Surprisal Theory is that the latter do not explain the greater processing difficulty of the intermediate embedded clause compared to the most embedded clause in the object double-center embedded relatives. These results speak in favor of a direct

role of the syntactic representation at the level of grammatical knowledge during sentence processing. Namely, from a hierarchical perspective, the primary source of difficulty in sentence processing does not seem to be only the presence of gaps but also the intervention of elements with relevant morphosyntactic features in long-distance dependencies.

Still, there is some evidence that the presence of a gap introduces a processing load even in a very short dependency. In Rel_Subj sentences such as (28), the question concerning the most embedded verb ('Who is chased? The little girl') elicited a better performance than the question concerning the intermediate embedded verb ('Who is indicated? The boys'). In this case, the facilitation extended to response times, which were faster when accuracy was higher. This pattern can be explained under the natural assumption that a filler gap dependency introduces a cost since the noun 'boys' is involved in a (short) filler gap dependency. In contrast, the noun 'little girl' is involved in no dependency.

Having said that, we must notice that the preference for the answer 'the little girl' over the answer 'the boys' in (28) might also be a recency effect due to the order of presentation of the two NPs in these sentences. Therefore, an explanation in terms of linear order cannot be entirely excluded for this specific case.

Moreover, the role of linear factors is also consistent with the pattern observed in control sentences; namely, questions concerning the last verb had higher accuracy and faster RTs than questions concerning the first and the second verb of the sentence (even if in Control sentences, this was expected also considering that questions concerning the last verb were composed of less words than the questions concerning the first and the second verb). This performance is determined by the position of the NP, which is the correct answer to the question: the most recent NP (the last one in the sentence) is associated to the best performance and the penultimate NP elicits an intermediate performance. In contrast, the first NP in the sentence is associated with the worst performance, and this can naturally be explained as a recency effect.

Therefore, we do not deny that linear factors play some role. They do so in control sentences and might do so in case of questions concerning the embedded verbs in subject double-center embedding configurations (as we just said, both the structural and the linear explanation can do the job). However, the linear explanation account faces a significant challenge from questions addressing the main verb in double-center embedding since it makes precisely the opposite prediction with respect to what is observed. Indeed, we observed that the most distant NP in linear terms is more easily retrieved. Therefore, we have clear evidence that structural factors (particularly hierarchical intervention) play a crucial role in the elaboration of long distance dependencies. In this case, the main source of difficulty in sentence processing seems to be the intervention of elements with morphosyntactic features such as lexical restriction. However, linear factors could play a role in other cases, for example, coordination of full clauses with most recent clause being better recovered than least recent clauses.

In conclusion, our data suggest that the underlying structure of the sentence, typically represented by a syntactic tree, is consulted in sentence processing. We are aware that the methodology used in this study allowed us to obtain only indirect evidence. In the future, we aim to obtain more direct evidence using a methodology that allows investigating online sentence processing.

5. Conclusions

The main aim of the present study was to compare different accounts of sentence complexity to achieve a better understanding of the factors involved. For this purpose, we investigated the processing of a configuration in which these models make different predictions, namely double-center embedding relatives. The results obtained are in line with the predictions of featural Relativized Minimality and cannot be fully explained by other accounts of sentence complexity. Therefore, structural factors seem to play a role in sentence complexity, and the main source of difficulty, at least when the sentences have a sufficient degree of complexity (e.g., when they contain subordinate structures), seems to be the intervention of elements with morphosyntactic features such as lexical restriction.

Data Accessibility Statement

All materials, the complete dataset, and the analyses scripts are fully available in the OSF repository of the current project (https://osf.io/7zg6m/).

Ethics and consent

The study was approved by the local ethical committee. Participants were treated in accordance with the ethical principles stated in the Declaration of Helsinki and given their written consent prior to the study participation.

Funding information

Giulia Bettelli was supported by a Fondazione Fratelli Confalonieri scholarship for PhD students.

Acknowledgements

We thank Carlo Toneatto for the technical support.

Competing interests

The authors declare that there is no conflict of interest. The authors have no relevant financial or non-financial interests to disclose.

Authors' contributions

Marco Sala: collected the data, drafted Section 1 (introduction) and Section 4 (general discussion). Giulia Bettelli: performed the analysis, drafted Section 3 (results). Beatrice Giustolisi: contributed to the initial design of the experiment, drafted Section 2 (methods). Alessandra Vergallito: contributed to the design of the experiment, performed the analysis and contributed to Section 2 and 3. Leonor Josefina Romero Lauro: contributed to the initial design of the experiment. Carlo Cecchetto: coordinated the teamwork and contributed to Sections 1 and 4. All authors provided critical feedback during all phases of the work and read and approved the final version

of the manuscript.

References

Baayen, Rolf H. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University press. https://doi.org/10.1017/CBO9780511801686

Baayen, Rolf H., Davidson, Dough J., & Bates, Douglas 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

Bates, Douglas, Mächler, Martin, Bolker, Ben, & Walker, Steve. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67(1),* 1–48. https://doi.org/10.48550/arXiv.1406.5823

Benjamini, Yoav, & Hochberg, Yosef. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological), 57(1),* 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In John R. Hayes (ed.), *Cognition and the Development of Language*, 279-362. New York: Wiley & Sons.

Cecchetto, Carlo, & Donati, Caterina (2015). *(Re)labeling, Linguistic Inquiry Monograph*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262028721.001.0001

Chomsky, Noam, & Miller, George A. 1963. Introduction to the Formal Analysis of Natural Languages. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (ed.), *Handbook of Mathematical Psychology, 269- 321*. New York: John Wiley & Sons. https://doi.org/10.2307/2269904

De Rosario-Martinez, Helios. 2015. PHIA. Post-hoc interaction analysis (R package v0.2-1). *CRAN: The R Foundation for Statistical Computing.* https://cran.r-project.org/web/packages/phia/index.html

Frazier, Lyn. 1987. Syntactic processing: evidence from Dutch. *Natural Language & Linguistic Theory, 5(4),* 519-559. http://doi.org/10.1007/Bf00138988

Frazier, Lyn, & Clifton, Charles J. 1989. Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes, 4,* 93–126. http://doi.org/10.1080/01690968908406359

Frazier, Lyn, & Flores d'Arcais, Giovanni B. 1989. Filler-driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language, 28,* 331–344. http://doi.org/10.1016/0749-596X(89)90037-5

Friedmann, Naama, Belletti, Adriana, & Rizzi, Luigi. 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua, 119,* 67-88. https://doi.org/10.1016/j.lingua.2008.09.002

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68,* 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gibson, Edward. 2000. The Dependency Locality Theory: A Distance-Based Theory of Linguistic Complexity. In Alec P. Marantz, Yasushi Miyashita, & Wayne O'Neil (ed.), *Image, language, brain: Papers from the first mind articulation project symposium, 94–126*. MA: MIT Press. https://doi.org/10.7551/mitpress/3654.001.0001

Gibson, Edward, & Iris Wu, Hsiao-Hung. 2013. Processing Chinese relative clauses in context. *Language and Cognitive Processes, 28:1-2,* 125-155. http://dx.doi.org/10.1080/01690965.2010.536656

Gordon, Peter C., Hendrick, Randall, & Johnson, Marcus 2001. Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(6),* 1411–1423. https://doi.org/10.1037/0278-7393.27.6.1411

Gordon, Peter C., Hendrick, Randall, & Johnson, Marcus. 2004. Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51(1),* 97–114. https://doi.org/10.1016/j.jml.2004.02.003

Grodner, Daniel, & Gibson, Edward. 2005. Consequences of the serial nature of linguistic input. *Cognitive Science, 29(2),* 261–290. https://doi.org/10.1207/s15516709cog0000_7

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics.*

Hale, John. 2003. The information conveyed by words in sentences. *Journal of Psycholinguistic Research, 32,* 101–123. https://doi.org/10.1023/a:1022492123056

Hsiao, Franny, & Gibson, Edward. 2003. Processing relative clauses in Chinese. *Cognition, 90,* 3–27. https://doi.org/10.1016/S0010-0277(03)00124-0

Huang, C.-T. James, Li, Y.-H. Audrey, & Li, Yafei. 2009. *The syntax of Chinese.* Cambridge: Cambridge University Press

Jäger, Lena, Chen, Zhong, Li, Qiang, Lin, C.-J. Charles, & Vasishth, Shravan. 2015. The subject-relative advantage in Chinese: Evidence for expectation-based processing.

Journal of Memory and Language, 79, 97–120. https://doi.org/10.1016/j.jml.2014.10.005

Karlsson, Fred. 2007. Constraints on multiple center-embedding of clauses. *Journal of Linguistics, 43(2)*, 365-392. https://doi.org/10.1017/S0022226707004616

King, Jonathan, & Just, Marcel A. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30(5)*, 580–602. https://doi.org/10.1016/0749-596X(91)90027-H

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition, 106*, 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Miyamoto, Edson T., & Nakamura, Michiko. 2013. Unmet expectations in the comprehension of relative clauses in Japanese. In *Proceedings of the Annual Meeting of the Cognitive Science Society, 35*, 3074-3079. https://escholarship.org/uc/item/1pp1w93k

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Reinhart, Tanya M. 1976. *The syntactic domain of anaphora.* Doctoral dissertation, Massachusetts Institute of Technology.

Rizzi, Luigi. 1990. *Relativized minimality.* Cambridge, MA: MIT Press. https://doi.org/10.9793/elsj1984.8.154

Rizzi, Luigi. 1997. The Fine Structure of the Left Periphery. In: Haegeman, Liliane (ed.) *Elements of Grammar.* Kluwer International Handbooks of Linguistics. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-5420-8_7

Ross, John R. 1967. *Constraints on variables in syntax*. Ph.D. Thesis, Cambridge: Massachusetts Institute of Technology.

Staub, Adrian, Foppolo, Francesca, Donati, Caterina, & Cecchetto, Carlo. 2018. Relative clause avoidance: Evidence for a structural parsing principle. *Journal of Memory and Language, 98*, 26 - 44, https://doi.org/10.1016/j.jml.2017.09.003

Traxler, Matthew J., Morris, Robin K., & Seely, Rachel E. 2002. Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language, 47(1)*, 69–90. https://doi.org/10.1006/jmla.2001.2836

Vasishth, Shravan, Chen, Zhong, Li, Qiang, & Guo, Gueilan. 2013. Processing Chinese Relative Clauses: Evidence for the Subject-Relative Advantage. *PLoS ONE 8(10):* e77006. https://doi.org/10.1371/journal.pone.0077006

Voeten, Cesko C. 2022. buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression. R package version 2.4. https://CRAN.R-project.org/package=buildmer

Yun, Jiwon, Whitman, John, & Hale, John. 2010. Subject-object asymmetries in Korean sentence comprehension. In Stellan Ohlsson and Richard Catrambone (ed.), *Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32).* https://escholarship.org/uc/item/4j65p53z