COGNITIVE SCIENCE
A Multidisciplinary Journal

How Children Process Reduced Forms: A Computational Cognitive Modeling Approach to Pronoun Processing in Discourse

Margreet Vogelzang,[a,b,*] 🆔 Maria Teresa Guasti,[c] Hedderik vanRijn,[d] Petra Hendriks[e]

[a]*Institute of Dutch Studies, University of Oldenburg*
[b]*Cluster of Excellence "Hearing4all", University of Oldenburg*
[c]*Psychology Department, University of Milano-Bicocca*
[d]*Departments of Experimental Psychology & Statistical Methods and Psychometrics, University of Groningen*
[e]*Center for Language and Cognition Groningen, University of Groningen*

Abstract

Reduced forms such as the pronoun *he* provide little information about their intended meaning compared to more elaborate descriptions such as *the lead singer of Coldplay*. Listeners must therefore use contextual information to recover their meaning. Across languages, there appears to be a trade-off between the informativity of a form and the prominence of its referent. For example, Italian adults generally interpret informationally empty null pronouns as in the sentence *Corre* (meaning "He/She/It runs") as referring to the most prominent referent in the discourse, and more informative overt pronouns (e.g., *lui* in *Lui corre*, "He runs") as referring to less prominent referents. Although children acquiring Italian are known to experience difficulties interpreting pronouns, it is unclear how they acquire this division of pragmatic labor between null and overt subject pronouns, and how this relates to the development of their cognitive capacities. Here we show that cognitive development can account for the general interpretation patterns displayed by Italian-speaking children and adults. Using experimental studies and computational simulations in a framework modeling bounded-rational behavior, we argue that null pronoun interpretation is influenced by working memory capacity and thus appears to depend on discourse context, whereas overt pronoun interpretation is influenced by processing speed, suggesting that listeners must

Correspondence should be sent to Margreet Vogelzang, Department of Theoretical and Applied Linguistics, University of Cambridge, 9 West Road, Cambridge CB3 9DP, UK. E-mail: mv498@cam.ac.uk

*Present addresss: Department of Theoretical and Applied Linguistics, University of Cambridge.

reason about the speaker's choices. Our results demonstrate that cognitive capacities may constrain the acquisition of linguistic forms and their meanings in various ways. The novel predictions generated by the computational simulations point out several directions for future research.

1. Introduction

When communicating, speakers tend to refer to entities or events in the world around them. To refer to a particular entity, speakers can often choose between various forms that differ in the amount of information they convey about their intended referent. For example, pronouns such as *he* provide little information about the intended referent compared to more elaborate descriptions such as *the lead singer of Coldplay*. Hence, pronouns can be said to be reduced forms. To determine the intended referent of a pronoun (i.e., its antecedent), listeners must use additional information such as context. Previous research has identified various contextual factors that influence the interpretation of pronouns, for example, syntactic function of the antecedent, structure of the global and local discourse, and frequency and recency of the antecedent (Ariel, 1990; Arnold, 1998; Givón, 1983; Gordon & Scearce, 1995; Kaiser, 2011; Kehler, Kertz, Rohde, & Elman, 2008). In addition, it has been suggested that the interpretation of pronouns is influenced by cognitive limitations such as working memory (WM) capacity (e.g., Almor, Kempler, MacDonald, Andersen, & Tyler, 1999; Serratrice, Sorace, Filiaci, & Baldo, 2012; Van Rij, Van Rijn, & Hendriks, 2013). However, little is known about how linguistic and contextual factors interact with these cognitive limitations when listeners process pronouns.

In contrast to English, Italian allows for two types of pronouns, namely overt pronouns (which are reduced compared to full NPs) and fully reduced null pronouns (which are reduced compared to overt pronouns). Although sentential subjects in Italian can be left unpronounced and are then called null pronouns, their meaning is included in the interpretation of the sentence. As a result, a sentence such as *Corre* "Runs" can be interpreted as "He/She/It runs." Across languages, there appears to be a trade-off between how informative a form is and how prominent its referent is. Informationally empty null pronouns, for example, are generally interpreted by Italian adults as referring to the most prominent referent in the discourse, whereas the more informative overt pronouns are interpreted as referring to less prominent referents (Carminati, 2002; Sorace & Filiaci, 2006). Thus, a division of pragmatic labor (Horn, 1984) appears to be present between the two reduced forms and their interpretations, with null pronouns being the unmarked or default form receiving an unmarked or predictable interpretation (namely reference to the most prominent referent) and overt pronouns signaling reference to a different, less prominent referent. Crucially, we hypothesize that this implies that the meaning of an overt pronoun is derived through an inference: If an overt pronoun rather than a null pronoun is used by the speaker, this signals to the listener that the speaker wanted to express a different

meaning than the unmarked meaning associated with the null pronoun (Gundel, Bassene, Gordon, Humnick, & Khalfaoui, 2010; Gundel, Hedberg, & Zacharski, 1993; Vogelzang, Foppolo, Guasti, Van Rijn, & Hendriks, 2020). As a consequence, we expect overt and null pronouns to not only have different distributions and interpretations, but to also be processed differently.

This hypothesized division of pragmatic labor between null versus overt pronouns in Italian will have to be acquired by children. Although Italian children have been found to show difficulty interpreting pronouns in an adult-like manner (Serratrice, 2007), it is unclear how they learn to use and interpret overt and null pronouns. Children's acquisition may be influenced by their limited and still developing cognitive capacities. In studies investigating children's linguistic performance by comparing different age groups, children have been shown to experience difficulty interpreting pronouns (e.g., English: Arnold, Brown-Schmidt, & Trueswell, 2007; Dutch: Hendriks, Koster, & Hoeks, 2014; Italian: Serratrice, 2007). One reason for children's difficulty could be that sufficient WM capacity is needed to keep track of the different referents in the discourse and their relative prominence (Van Rij et al., 2013; Vogels, Krahmer, & Maes, 2015). In addition, listeners need to reason about the choices of the speaker (e.g., Brown-Schmidt & Hanna, 2011; Frank & Goodman, 2012; Goodman & Frank, 2016; Gundel et al., 1993, 2010; Hendriks & Spenader, 2006; Nadig & Sedivy, 2002; Vogelzang et al., 2020). Van Rij, Van Rijn, and Hendriks (2010) argue that in order to consider the perspective of the speaker, the listener needs sufficient processing speed. So WM capacity and processing speed have independently been argued to affect pronoun processing in discourse. Therefore, we focus on these two cognitive factors in this study.

Addressing the question how children acquire the division of pragmatic labor between two reduced forms with the same grammatical function, namely null and overt subject pronouns in Italian, and how this relates to the development of their cognitive capacities, in particular WM capacity and processing speed, we present a computational cognitive model (from here on: cognitive model) that describes in detail what processes may underlie pronoun interpretation. Implementing the model based on current theoretical insights about pronoun processing while abiding by the cognitive constraints provided by a cognitive architecture will result in a bounded-rational model (see Vogelzang et al., 2017), which contrasts with purely rational models proceeding from ideal listeners and speakers with unlimited processing capacity (Frank & Goodman, 2012; Goodman & Frank, 2016). The model will be implemented in the cognitive architecture ACT-R (Anderson, 2007), which pre-specifies cognitive constraints such as the time required to retrieve information from memory, or the number of visual features that can be attended at once. By implementing a pronoun processing model in this architecture, we can assess the influence of, for example, WM capacity on pronoun processing through the manipulation of cognitive constraints (see also Van Rij et al., 2013). As an additional benefit, we can examine the interplay between different factors, something that is difficult to study through psycholinguistic experiments. The output of the model can directly be compared to behavioral data of human participants. This allows for the explicit testing of hypotheses about the role of cognitive capacities in language processing and development.

The paper is organized as follows. First, we present an ACT-R model simulating the processing of pronouns by Italian adults, based on existing data. On the basis of this adult model, which specifies the effects of WM capacity and processing speed, we formulate several hypotheses about children's processing of pronouns. We present a new behavioral experiment with children that tests these hypotheses. Next, we use the same cognitive model to simulate the patterns observed in this experiment. By comparing these child simulations with the adult simulations, this study sheds more light on the factors influencing children's development and processing of reduced forms.

2. Experimental data on adults' processing of pronouns

To inform our pronoun processing model, we used the data of Vogelzang et al. (2020). With a referent selection task measuring pupillary responses, they examined native Italian adults' interpretation of three types of anaphoric subjects: full noun phrases (NPs) such as *the hedgehog*, the overt pronouns *lui* ("he") and *lei* ("she"), and null pronouns (Ø). A sample story illustrating the presentation of the three different anaphoric subjects in a discourse is shown in (1).

(1)	1.	Il riccio	compra	della moquette	per il soggiorno.
		The hedgehog buys		*some carpet*	*for the living-room.*
		The hedgehog is buying		*some carpet*	*for the living room.*
	2.	Ieri	il riccio	ha raccontato	al topo una storia,
		Yesterday	*the hedgehog*	*has told*	*to-the mouse a story,*
		Yesterday	*the hedgehog*	*told*	*the mouse a story,*
	3.	mentre	il riccio/lui/Ø	si annoiava	davanti alla tv.
		while	*the hedgehog/he/Ø*	*himself bored*	*in-front to-the TV.*
		while	*the hedgehog/he/Ø*	*was bored*	*in front of the TV.*

To test participants' interpretation of the critical anaphoric subject in the third clause of the discourse, which could be one of two referents, a question about this subject was asked (e.g., *Who was bored?*). Both referents were displayed as pictures on the computer screen and participants were asked to select one of these referents as the answer to the question by pressing the left or right button, corresponding to the left or right picture on the screen. In total, 48 critical stories and questions were auditorily presented to participants. In each story, the referent that is mentioned first (e.g., *the hedgehog* in (1)) is the most prominent referent in the discourse and hence is the discourse topic. This referent is mentioned as the subject twice to give it a high prominence. The second referent (*the mouse* in (1)) is less prominent and is a non-topical referent. Vogelzang et al.'s original experiment additionally tested the interpretation of anaphoric objects by also including 48 items with questions about the object, but those items will be treated as filler items here.

The results of Vogelzang et al. (Fig. 1) show that full NPs, which served as an unambiguous baseline condition, were almost always interpreted correctly by adults. Null
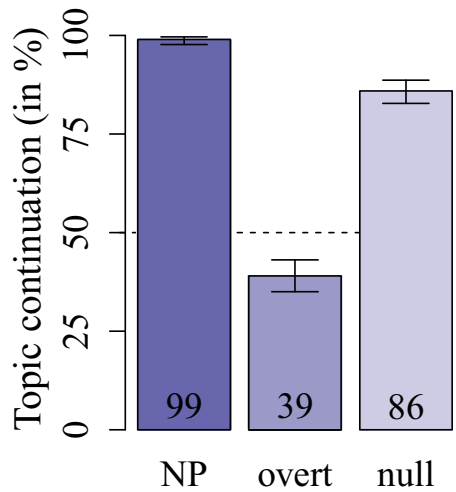
Fig. 1. Experimental data from Vogelzang et al. (2020; reprinted with permission), showing the percentage of adults' responses indicating a topic continuation for the three types of anaphoric subject in Italian (NP: full NP; overt: overt pronoun; null: null pronoun).

pronouns were most often interpreted as referring to the topic (86% of cases). In contrast, Italian adults' interpretations of overt pronouns varied, as overt pronouns were most frequently interpreted as referring to the non-topical referent (61% of cases) but also often as referring to the discourse topic (39% of cases).

In addition to offline interpretations, Vogelzang and colleagues measured pupil dilation as an indication of cognitive effort during language processing. Overt pronouns were found to be more effortful to process than null pronouns, which in turn were found to be more effortful to process than unambiguous full NPs.

Based on these findings, we formulate two hypotheses that we will test with a cognitive model. First, the offline results show that null pronouns are mostly interpreted as referring to the discourse topic, indicating that adults have little problems processing the discourse and keeping track of the discourse referents and their prominence. Therefore, we hypothesize that the variation in adults' interpretation of overt pronouns is not due to discourse factors, but to linguistic or cognitive constraints (hypothesis 1).

Second, the pupil dilation measures show that, for adults, overt pronouns are more effortful to process than null pronouns. We hypothesize that this is because in Italian the interpretation of overt pronouns, but not of null pronouns, requires perspective taking (hypothesis 2). Perspective taking entails that when interpreting a potentially ambiguous form, such as a pronoun, listeners can eliminate the ambiguity by taking into account the alternative forms the speaker could have used but did not use. This process will be explained in more detail in the next section, followed by a description of the implementation of the cognitive model.

3. Modeling study 1: Adults' processing of pronouns

This section presents a cognitive model of Italian pronoun processing, which will be used to simulate adult performance. The model incorporates linguistic constraints as well as a linguistic mechanism of perspective taking to account for the processing and interpretation of pronouns.

3.1. Perspective taking

We implemented the linguistic component of pronoun processing in terms of the constraint-based linguistic framework optimality theory (OT; Prince & Smolensky, 2004; for earlier cognitive models based on this approach, see Hendriks, Van Rijn, & Valkenier, 2007; Misker & Anderson, 2003; Van Rij et al., 2010). OT accounts for the interaction between linguistic constraints through its mechanism of optimization over potential forms or meanings. In addition, OT accounts for perspective taking in language use through bidirectional optimization (Blutner, 2000).

Optimality theory distinguishes the production perspective of the speaker from the comprehension perspective of the listener. In OT production, an input meaning is mapped onto potential forms for expressing that meaning. This reflects the view that speakers can potentially express a certain meaning in multiple ways by using different forms. On the basis of a set of hierarchically ranked constraints that make up the grammar, the form that satisfies the constraints best from a set of potential forms is selected as the optimal output. Crucially, constraints in OT may conflict. In case of a conflict, a higher ranked constraint has priority over a lower ranked constraint. The produced output form is, according to the grammar, the best way to satisfy these conflicting constraints and hence the optimal way to convey the intended meaning. Likewise, in OT comprehension, the optimal meaning for a given input form is the meaning from a set of potential meanings that satisfies the same constraints best. In other words, listeners interpret a given form as conveying the meaning that, according to the grammar, is the optimal way to interpret that form. The production perspective and the comprehension perspective are combined in bidirectional optimization, which can be seen as a formalization of perspective taking. When a listener considers potential meanings for a given form, the purpose of perspective taking is to also include the choices of the speaker as a factor in deciding on the optimal meaning. This reflects the view that listeners reason about the forms the speaker used or could have used (cf. De Hoop & Kramer, 2006; Hendriks & Spenader, 2006). The effect of perspective taking is that certain meanings are ruled out for a given form, because the speaker would not have used that form to express those meanings.

Bidirectional optimization can be operationalized as a two-step procedure, starting with an optimization step from the adopted perspective, followed by an optimization step from the opposite perspective (Hendriks et al., 2007; Van Rij et al., 2013). In pronoun interpretation, listeners thus start from the comprehension perspective. In the first step, they determine the optimal interpretation for the encountered pronoun by generating potential interpretations for this pronoun and applying constraints on pronoun interpretation. In the

second step, listeners take the production perspective in order to determine if a speaker would indeed have used the encountered pronoun for expressing the selected meaning. If the optimal form from the production perspective is identical to the encountered form, the selected meaning is bidirectionally optimal. If, on the other hand, the optimal form from the production perspective is different from the encountered form, the selected meaning is not bidirectionally optimal and another meaning must be selected.

Based on these optimization processes, the optimal interpretation of a null pronoun in Italian is reference to the discourse topic. In contrast, the optimal interpretation of an overt pronoun in Italian cannot be reference to the topic, because the second step in bidirectional optimization reveals that the optimal way for a speaker to express reference to the topic is by using a null pronoun. Hence, an overt pronoun must refer to another referent. This perspective-taking approach to overt pronouns in Italian is in line with linguistic research on the use and interpretation of referring expressions in discourse, in particular by Gundel and colleagues (e.g., Gundel et al., 1993, 2010). Gundel and colleagues argue that referring expressions can be placed in an implicational hierarchy reflecting the assumed cognitive status of their referent (see Ariel, 1990; Givón, 1983, for alternative referential hierarchies). Listeners straightforwardly interpret referring expressions that are at the high end of this so-called givenness hierarchy, such as null pronouns in Italian, but must perform a pragmatic inference to interpret referring expressions that are lower in the hierarchy, such as overt pronouns in Italian.

3.2. The cognitive architecture ACT-R

Our cognitive model[1] is implemented in the computational cognitive architecture ACT-R (Anderson, 2007). ACT-R includes assumptions about human cognition based on psychological experiments and theories, and thus provides a psychologically realistic modeling framework. Models developed within this cognitive architecture share these assumptions in the form of predefined general processes and parameter values, which are therefore not fit to a specific dataset. Because of the shared cognitive assumptions, the outcomes of cognitive models developed within such a framework are intended to be psychologically plausible. This type of modeling falls within a modeling tradition in which general principles of cognition are used to explain human behavior on specific tasks (e.g., Daily, Lovett, & Reder, 2001; Taatgen & Anderson, 2002; Van Rij et al., 2013; Van Rijn, Van Someren, & Van der Maas, 2003).

Of the cognitive assumptions incorporated within ACT-R, the ones about memory storage and retrieval are of particular importance for the implementation of the linguistic aspects of the model, as they guide the retrieval of linguistic forms and meanings and linguistic constraints. By manipulating these mechanisms, we can simulate changes in cognitive capacities, and hereby assess the influence of these capacities on linguistic processing. This will enable us to investigate the intricate interplay between discourse, linguistic, and cognitive factors in pronoun processing.

ACT-R consists of several functional modules. These modules interact to allow for complex processing, such as linguistic processing. The most important modules for the

current model are the declarative memory module and the procedural memory module. Factual information (e.g., *a hedgehog is an animal*) in declarative memory is represented as chunks. Every chunk has an associated activation, reflecting an estimate of the usefulness of that chunk in the current context based on past experiences. A chunk can only be retrieved from memory if its activation is sufficiently high. When multiple chunks are competing to be retrieved, the chunk with the highest activation will be retrieved. A chunk's activation also determines the time required for its retrieval: the higher its activation, the faster its retrieval. The procedural memory module represents knowledge about how to perform actions, represented as production rules. If the conditions of a rule are met, the rule can be selected and its actions (e.g., initiate a retrieval from memory) will be executed. Production rules are selected and executed serially, and execution takes time. Performing a task such as pronoun processing requires the usage of both the declarative memory module, in which referents encountered in the discourse are stored, and the procedural memory module, which regulates the retrieval and use of these referents.

Using the cognitive architecture ACT-R, human behavior observed in experiments can be modeled by simulating a number of participants performing a specific task. The task of our model is to process a story and for each encountered pronoun determine its referent. The model builds on earlier work by Hendriks et al. (2007), Van Rij et al. (2010), and Van Rij et al. (2013) on Dutch. As such, very few additional assumptions need to be made in our model; large parts of the implementation have been used in other models to explain other datasets. Thus, rather than fitting our model to a specific dataset, the processes and parameter values of the ACT-R architecture and existing ACT-R models are reused, in particular of models by Anderson (2007), Hendriks et al. (2007), Van Rij et al. (2010), and Van Rij et al. (2013).

3.3. Methods

The model was designed to process stories in Italian. The subject form in the final clause of the story is either a full NP, an overt pronoun, or a null pronoun. In each run (simulating one participant), the model is presented with training items first, before proceeding to the test phase. In the test phase, 16 stories like (1) are presented per subject form, resulting in 48 stories in total. No feedback is provided during the training or test phase.

3.4. Implementation

The model consists of a discourse processing component and a sentence processing component. In the following sections, we will describe these two components and how they are influenced by WM capacity and processing speed.

3.4.1. Discourse processing

Before the sentence with the pronoun is processed (which is explained in Section 3.4.2), the model processes the preceding discourse by keeping track of all the referents in the story (based on Van Rij et al., 2013). Each referent, represented as a chunk,

has an activation, which is the sum of its base-level activation, its spreading activation, and a noise component. Referent activation is used as a measure of discourse prominence. Based on default mechanisms of ACT-R, the base-level activation of a referent is influenced by its recency and frequency in the discourse. This activation decays over time, but is increased when the referent is encountered again, and is calculated according to the following equation:

$$\text{(a)} \quad B_i = \ln(n/(1-d)) - d \times \ln(L).$$

In this formula, B_i is the base-level activation of chunk i, d is the decay parameter which has a default value of 0.5, n is the number of presentations of chunk i, and L is the time since chunk i was created. Thus, the formula describes that the base-level activation of a chunk is determined by the number of repetitions of the chunk and the time since its creation: Chunks that are encountered more often and/or more recently will have a higher base-level activation.

Added to a chunk's base-level activation is the spreading activation it receives; a chunk that is retrieved from memory will also activate chucks that are associated with it. Spreading activation is determined according to the following simplified formula (cf. Anderson, 2007):

$$\text{(b)} \quad S_i = \sum_j W S_{ji}.$$

In this formula, the spreading activation to chunk i is determined by the strength of association S between all chunks j and chunk i, modulated by the amount of spreading activation W. This formula indicates that a chunk that is associated with other chunks will receive additional activation from these other chunks, the amount of which depends on the strength of the associations between the chunks, so that chunks with many strong associations with other chunks will have a higher activation.

Besides the base-level activation and spreading activation, a chunk's activation at retrieval is also influenced by noise (see Anderson, 2007, for more details). Noise on activation can cause a chunk with a lower total activation to be retrieved over a chunk with a higher total activation. In the context of our study, this would mean that a less prominent referent may occasionally be retrieved as the discourse topic, instead of the most prominent referent.

3.4.1.1. WM capacity: In the model, keeping track of the discourse referents is influenced by WM. ACT-R has no separate WM component (see, e.g., Borst, Taatgen, & Van Rijn, 2010), but explains WM effects as variations in activation patterns in declarative memory. Our model uses an implementation of WM based on spreading activation to keep discourse referents active in memory (based on the model of Daily et al., 2001, and also applied in Van Rij et al., 2013). Specifically, we assume that activation is spread by the grammatical subject to all discourse referents associated with that subject. This means

that when a referent is referred to as the grammatical subject of a clause, activation will spread to the chunk in declarative memory that represents this referent. Similarly, if a subject pronoun is resolved as referring to a specific referent, the chunk in declarative memory that represents this referent will then receive spreading activation because of its association with the subject pronoun. As a consequence, the referent associated with the subject tends to be the most strongly activated referent and hence will likely be identified as the discourse topic.

The amount of spreading activation reflects WM capacity. For adults, WM capacity will generally be high, resulting in a boost of spreading activation to the referents associated with the grammatical subject. In line with Anderson, Reder, and Lebiere (1996) and Van Rij et al. (2013), we set the spreading activation parameter W to a default value of 1.0 to simulate high WM capacity.

Every story in the study has a similar structure. In the first clause, the first referent (*the hedgehog* in (1)) is introduced as the grammatical subject. In the second clause, the same referent is repeated in subject position, and a second referent (*the mouse* in (1)) is introduced in object position. In the third and final clause, one of three anaphoric subjects (full NP, overt pronoun, null pronoun) is used. At the start of clause 3, *the hedgehog* is the most likely local discourse topic based on frequency and subjecthood in clauses 1 and 2. If the pronoun in clause 3 is interpreted as referring to *the hedgehog*, this is thus a continuation of the topic. If, in contrast, the pronoun is interpreted as referring to *the mouse*, this indicates a topic shift from *the hedgehog* to *the mouse*.

At any point in the story, the discourse topic is the referent with the highest activation in the model's memory at that time. Figs. 2 and 3 show the referent activation in memory over time of the two discourse referents when WM capacity is high, simulating adult processing. In Fig. 2 the pronoun is interpreted by the model as referring to *the hedgehog*, illustrating topic continuation. In Fig. 3 the pronoun is interpreted by the model as referring to *the mouse*, illustrating a topic shift.

3.4.2. Sentence processing

When presented with a story, the model builds a syntactic representation of each sentence in a word-by-word fashion. Every time a new word is encountered, lexical and syntactic information is retrieved from declarative memory. This information is used to attach the word to a syntactic structure. Thus, linguistic processing takes place in an incremental manner (cf. Lewis & Vasishth, 2005; Van Rij et al., 2013). Lexical information (e.g., word category, gender, animacy) is retrieved to assess whether a word could be a potential antecedent of the pronoun.

Processing a pronoun is part of sentence processing. We investigate the interpretation of three subject forms: full NPs, overt pronouns, and null pronouns. When a full NP is encountered, discourse processing and sentence processing take place, but no ambiguity resolution is needed, as the NP is unambiguous. Whenever an overt pronoun is encountered, it is resolved immediately (in line with Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Badecker & Straub, 2002; Vogelzang, Hendriks, & Van Rijn, 2016; but contra McDonald & MacWhinney, 1995; Stewart, Holler, & Kidd, 2007). Null pronouns
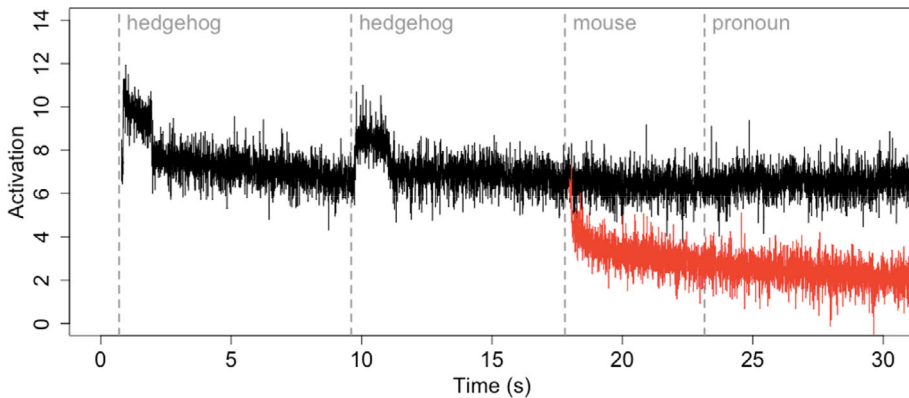
Fig. 2. Referent activation values in memory over time in a high WM capacity model, simulating an adult. The black line shows the activation of the firstly introduced referent (here, *the hedgehog*); the red line shows the activation of the other referent (here, *the mouse*). Because the firstly introduced referent remains the highest activated referent throughout the story, this simulation illustrates topic continuation.
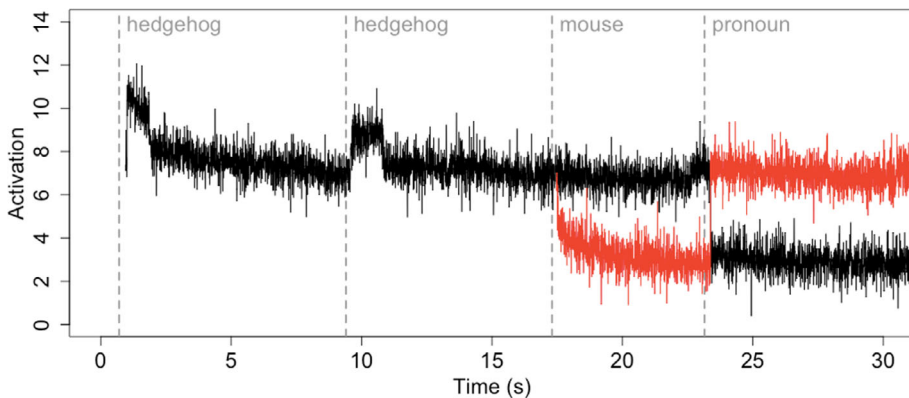

Fig. 3. Referent activation values in memory over time in a high WM capacity model, simulating an adult. The black line shows the activation of the firstly introduced referent (here, *the hedgehog*); the red line shows the activation of the other referent (here, *the mouse*). Because the firstly introduced referent is not the highest activated referent anymore after encountering the pronoun, this simulation illustrates a topic shift.

do not have an overt form, but when a finite verb is encountered without a preceding grammatical subject, the model assumes that a null pronoun was used and processes the sentence as such. If pronoun processing is not completed in time, it is aborted and the model continues by processing the next word.

Pronoun processing in the model is constrained by the following three hierarchically ordered linguistic constraints:

[1] Null pronouns refer to the topic
[2] Avoid full NPs
[3] Avoid overt pronouns

The first constraint restricts the use and interpretation of null pronouns, stating that they must refer to the most prominent referent at that point in the discourse, namely the current discourse topic (similar to overt pronouns in non-null subject languages like English, cf. Beaver, 2004; Grosz, Joshi, & Weinstein, 1995; Hendriks, Englert, Wubs, & Hoeks, 2008; Van Rij et al., 2013). No comparable constraint is used for overt pronouns, as we hypothesize that their interpretation is derived from the interpretation of null pronouns through perspective taking (see Section 3.1). The second and third constraints (adopted from Hendriks & Spenader, 2006; Van Rij et al., 2010) reflect the view that using linguistic material is costly and speakers prefer to be as economical as possible, providing as little information as possible. Thus, speakers prefer to use null pronouns over overt pronouns, and prefer to overt pronouns over full NPs.

Pronoun processing in the model takes place in four steps (Fig. 4).

When determining the discourse topic (Fig. 4, Box A, shown in detail in Fig. 5), the referent with the highest activation is retrieved from memory. Given sample story (1), this will usually be *the hedgehog*, but the non-topical referent *the mouse* can also occasionally be retrieved due to noise.

Once the discourse topic has been determined, the model interprets the pronoun from the perspective of the listener (Fig. 4, Box B, shown in detail in Fig. 6). All potential meanings of the pronoun (in the model: reference to the topic and reference to the non-topic) are iteratively retrieved as candidate meanings.

The model then retrieves a constraint and evaluates the two candidate meanings on the basis of this constraint. If one of the meanings violates the constraint, this meaning is discarded. The model continues this cycle until (a) there is only one candidate left, (b) there are no more constraints to be retrieved, or (c) the process is aborted due to time running out.

If the input of the interpretation step is a null pronoun, then reference to a non-topical antecedent violates constraint [1]. As a consequence, given sufficient time, reference to the topic is selected as the optimal meaning. Alternatively, if the input is an overt pronoun, constraint [1] does not restrict its interpretation, and neither do the other two constraints. Therefore, the overt pronoun remains ambiguous between reference to the topic and reference to the non-topical referent. The model will now randomly select one of these two meanings, leading to chance performance at this point in processing. Once an
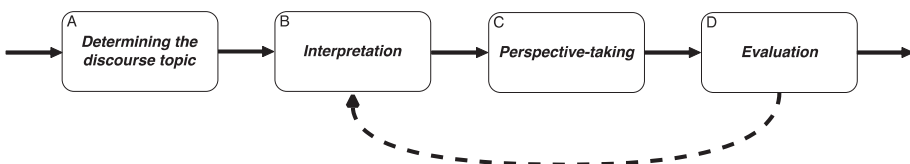

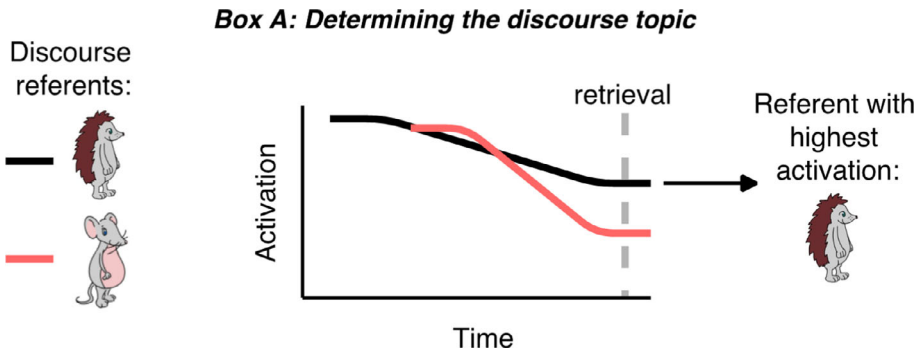Fig. 4. The four pronoun processing steps in the model.

Fig. 5. The discourse topic is determined by retrieving the referent with the highest activation from memory.
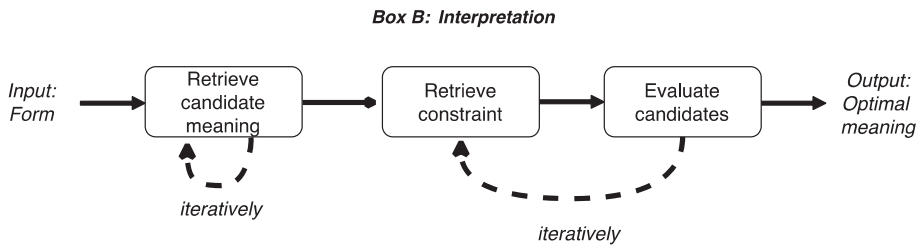

Fig. 6. In the interpretation step, the model first retrieves candidate meanings and then iteratively evaluates these meanings on the basis of retrieved constraints.

optimal meaning has been selected from the interpretation perspective, the model continues with the perspective-taking step.

In this step (Fig. 4, Box C, shown in detail in Fig. 7), the model takes the production perspective. The optimal meaning of the interpretation step is now the input, and the
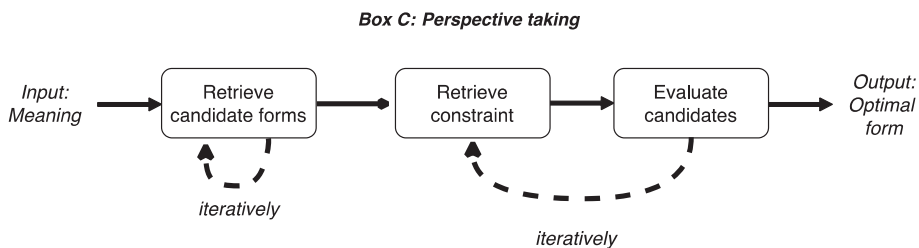

Fig. 7. In the perspective-taking step, the model takes the production perspective and retrieves candidate forms (full NP, overt pronoun, and null pronoun) for the optimal meaning selected in the previous step. Next, it iteratively evaluates these forms using the same procedure and constraints as in the interpretation step in Fig. 6. The only difference is that now the input is a meaning and the output is a form.

model determines the optimal form (i.e., a full NP, overt pronoun, or null pronoun) for this meaning. The same constraints are used as in the interpretation step.

If the input is reference to the topic, then constraint [1] does not restrict the use of any of the forms. The use of a full NP violates constraint [2], and the use of an overt pronoun violates constraint [3]. So, the optimal form to express reference to the topic is a null pronoun, which does not violate any of the constraints. Alternatively, if the input is reference to the non-topic, the use of a null pronoun violates constraint [1]. Again, the use of a full NP violates constraint [2] and the use of an overt pronoun violates constraint [3]. Because constraint [3] is the weakest, the optimal form to express reference to the non-topical referent is an overt pronoun.

In the final evaluation (Fig. 4, Box D, shown in detail in Fig. 8), this optimal form is compared to the original input. If the optimal form is identical to the original input, the model concludes that the selected optimal meaning is correct. This will be the case for null pronouns, as the optimal meaning for a null pronoun is the topic and the optimal form for a topic is again a null pronoun. The model will therefore select the discourse topic as the referent of the null pronoun. If, on the other hand, the optimal form is not identical to the original input, the model will block the initially selected optimal meaning and return to the interpretation step to select another meaning. In the case of an overt pronoun, a meaning is selected randomly in the interpretation step and this can thus be reference to the topic. In this case, the perspective-taking step will reveal that the optimal form for referring to the topic is a null pronoun. Thus, the input form of the interpretation step (an overt pronoun) will differ from the output form of the perspective-taking step (a null pronoun). Reference to the topic is therefore blocked and the interpretation of the overt pronoun will have to be revised.

3.4.2.1. Processing speed: Sentence processing is constrained by the time available between the presentation of subsequent words. In the model, a new word is presented approximately every second (see Section 7 for discussion). As the model needs to have
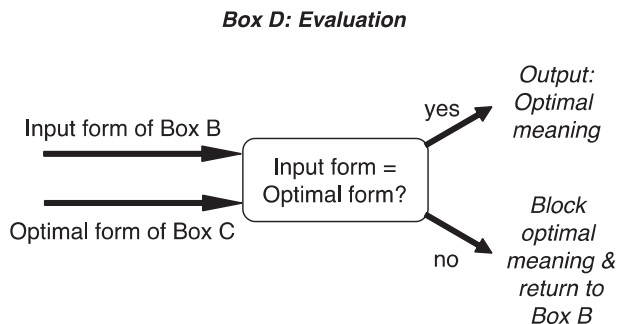

Fig. 8. In the evaluation step, the model compares the input of the interpretation step (Box B) to the optimal form of the perspective-taking step (Box C). If the two forms are identical, the optimal meaning from Box B is selected as the final interpretation. If the two forms differ, the optimal meaning from Box B is blocked and the model returns to Box B to select a different meaning.

processed a word before the next word can be processed, it requires sufficient processing speed.

Processing speed increases as a result of language experience. For adults, processing speed will generally be high because of extensive language experience, including specific experience with pronoun processing. Two standard ACT-R mechanisms are responsible for the model's increase in processing speed: activation and production compilation. When a particular linguistic form or meaning is encountered regularly and each time the relevant chunks are retrieved from memory, their activation increases and as a result they will be retrieved faster next time. Another way for the model to increase processing speed is through production compilation. Production compilation is a mechanism that allows the model to combine multiple production rules into one, so that tasks that have been performed frequently take fewer cognitive steps, and thus less time (Taatgen & Anderson, 2002). Moreover, production compilation can create new rules that include chunk information and therefore removes the need for time-consuming and potentially erroneous memory retrievals.

Language experience is provided to the model in the form of a training phase preceding the test phase. In the training phase, the model is trained on the processing of overt and null pronouns without a discourse (in a ratio of 30% overt pronouns and 70% null pronouns, reflecting the ratio of use of null and overt pronouns in Italian, see Lorusso, Caprin, & Guasti, 2005; Serratrice, 2005). This provides the model with experience in resolving pronouns without training it on specific discourses. During training, the model is instructed to interpret bare null and overt pronouns, through which it practices pronoun processing. This way, the model increases the efficiency of the production rules involved and the activation of the linguistic constraints that are relevant for pronoun resolution. The more training the model receives, the faster it will become. Importantly, all of this training consists of unsupervised learning, in the sense that the model receives no feedback on the optimal interpretation of a presented pronoun. The model learns to associate a form with a particular meaning solely based on the activation of the linguistic constraints and the successful completion of perspective taking (see Section 3.4.2). A plot of the model's development of the interpretation of null and overt pronouns with more training is presented in Fig. A1 in the Appendix. This plot shows that the number of overt pronouns interpreted as a topic continuation slowly decreases in the model with more training.

3.5. Procedure

The model described above was used to simulate the participants of Vogelzang et al. (2020). The model performed all steps that these participants also had to perform: hearing and processing a story with an anaphoric subject, being presented with a question asking for the interpretation of the anaphoric subject, and answering that question by choosing the antecedent.

We ran the model for 40 simulations, simulating 40 adult participants. In each simulation, the model was first trained on 2,000 items. These training items provided the model

with prior experience in resolving pronouns and increased its processing speed. The test phase consisted of 48 items with anaphoric subjects (null pronouns, overt pronouns, full NPs).

As noise influences the activation values of chunks, causing variations in retrieval during the training and test phase, every simulation will differ slightly. This way, running the model multiple times will result in a dataset representing variability not unlike the individual variation present in a group of human participants.

3.6. Results

The output of the model is shown in Fig. 9, along with the responses by the human participants in the study by Vogelzang et al. (2020).

The model accounts for the general patterns in the data (model fit Pearson r^2 = .99): Unambiguous full NPs are interpreted correctly, null pronouns often refer to the discourse topic, and overt pronouns vary in their interpretation but are mostly interpreted as referring to the non-topical referent.

3.7. Discussion

As the simulations show, our cognitive model accounts for the general pattern of interpretation of overt and null pronouns by Italian adults. Both language experience (implemented as training) and WM capacity (implemented as spreading activation) contribute to the model's performance, as becomes apparent when examining additional model simulations in which one of these components is absent. With high WM capacity but without training (Fig. A2 in the Appendix), overt pronouns are interpreted as a topic continuation
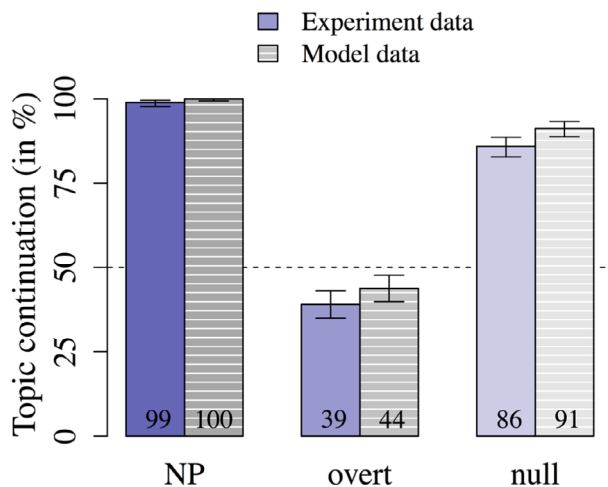

Fig. 9. Experimental data of adult participants (Vogelzang et al., 2020) and simulation data of the adult model on the interpretation of three types of anaphoric subjects in Italian (NP: full NP; overt: overt pronoun; null: null pronoun). Error bars are derived from logistic analysis.

around 50% of the time, indicating chance performance due to insufficient processing speed to complete perspective taking. Even after training, without WM aiding the identification of the discourse topic (Fig. A3 in the Appendix), null as well as overt pronouns are interpreted as a topic continuation only around 50% of the time. Thus, both sufficient training and sufficient WM capacity are needed to model adult performance on pronoun interpretation in Italian.

Two hypotheses, based on empirical data of human participants, were tested in the model. The first hypothesis was that adults show variation in their interpretation of overt pronouns because of linguistic and cognitive factors. This hypothesis was tested by providing the model with high WM capacity and high processing speed to simulate adult performance. The model results shown in Fig. 9 indicate that, even with high processing speed, the limited processing time inherent in ACT-R models causes occasional problems with perspective taking. This provides a viable explanation for Italian adults' variation in their interpretation of overt pronouns. The simulations furthermore show that this variation does not stem from difficulties with discourse processing, as the model is able to keep track of the discourse referents and their prominence if WM capacity is high (see Figs. 2 and 3).

The second hypothesis was that overt pronouns are derived from null pronouns through perspective taking and hence are more difficult to process than null pronouns. In the model, the interpretation of null pronouns is dependent on linguistic constraints, whereas the interpretation of overt pronouns is derived through perspective taking. The model results indeed show a clear interpretational preference for null pronouns, whereas the interpretation of overt pronouns varies more. This shows that the hypothesized difference between null and overt pronouns can account for the empirical data.

As predicted, adults' pronoun resolution was influenced by an interaction between linguistic and cognitive factors. Based on these findings, we can formulate novel predictions about how children acquire and process these reduced forms. First, because WM capacity is needed to keep track of the discourse referents, we predict that children, who generally have lower WM capacity (e.g., Gathercole, Pickering, Ambridge, & Wearing, 2004; Van Rijn et al., 2003), will have difficulty with discourse processing and keeping referents activated in memory. This will lead to mistakes in retrieving the referents of unambiguous full NPs and determining the discourse topic, resulting in interpretation errors for full NPs and fewer topical interpretations for null pronouns compared to adults. In other words, children may sometimes have difficulties remembering who did what to whom in the story, even when full NPs were used. Children are expected to become more adult-like in their interpretations of full NPs and null pronouns as they grow older and their WM capacity increases.

Second, children are expected to process linguistic information slower than adults because they have less language experience. As a consequence, children are expected to experience difficulty in perspective taking. This will influence the interpretation of overt pronouns, which requires perspective taking, and result in chance level performance. We tested these predictions in a pronoun interpretation experiment with children aged 6–8.

4. Experimental study 1: Children's processing of pronouns

In this section, we present data from an experiment with Italian children testing their interpretation of pronouns, using similar materials and procedure as in Vogelzang et al. (2020). This allows us to directly compare children's and adults' interpretations of pronouns.

4.1. Methods

4.1.1. Participants

Fifty-two children (28 girls, age 6;0–8;9, mean age 7;4) participated in the experiment. The children were recruited through a primary school in Milan, Italy. They were all native speakers of Italian and had normal or corrected-to-normal vision and hearing. The school's approval and parental consent were obtained prior to testing. Ethical approval was obtained from the University of Milano-Bicocca (prot. 20974/13). For the analyses, the children were split up into three age groups (see Table 1).

4.1.2. Materials

The same referent selection task was used as in the study with adults (Vogelzang et al., 2020), as these materials were explicitly designed to be appropriate for children, too. This means that we can compare the results of the children in this study to the results of the adults in the study of Vogelzang et al. (2020), who were tested on the same experimental items and filler items. An example story was already presented in (1).

Children's interpretation of three types of anaphoric subjects (full NPs, overt pronouns, null pronouns) was determined on the basis of referent selection questions such as *Chi si annoiava?* ("Who was bored?"). Both referents in the story were depicted on the computer screen from the start of the story until an answer was given (Fig. 10). To answer the question, the child could choose between the two referents.

4.1.3. Design and procedure

The experiment consisted of 30 auditorily presented critical stories and 30 filler stories,[2] divided over two test sessions. Before the first session, the children received a practice block of six stories. The sessions took about 20 min (session 1) and 15 min (session 2).

Table 1
Number of children (*n*) and mean age per age group

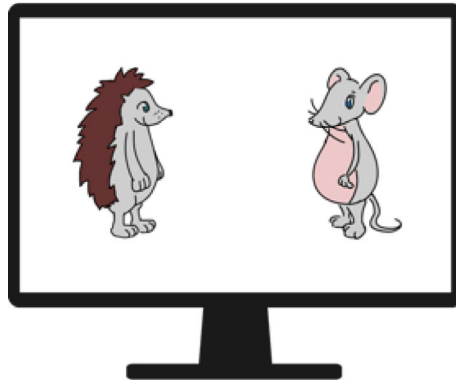| Group	*n*	Mean Age
6-year-olds	19	6;5
7-year-olds	20	7;6
8-year-olds	13	8;4

Fig. 10. Example of the presentation of two referents on the computer screen.

Anaphoric subject form and the location of the referents on the screen were counterbalanced. The order of presentation of the stories was randomized per participant.

The children were tested in a quiet room at school. They heard the stories and questions while looking at a computer screen, and they were instructed to listen to the stories carefully and to answer the questions by pressing the left or right trigger button on a gamepad, corresponding to the picture of the animal on the left or right of the screen, respectively.

4.2. Results

One of the 52 children only completed the first session of the experiment, so 1,545 critical trials were recorded in total. Seven trials (0.5%) in which the participant took longer than 20 s to respond were excluded, resulting in 1,538 remaining trials for analysis.

Whereas the adult participants heard 48 critical stories, the child participants were presented with a subset of 30 stories. To verify that this would not influence our results, we compared children's results on the 30 stories to adults' results on the same subset of 30 stories as well as to adults' results on the full set of 48 stories. Analyses indicated that the results were qualitatively similar. Therefore, the original adult dataset with 48 stories will be used when comparing the adult and child data.

Children's and adults' responses in the three subject conditions, presented as the percentage of responses indicating a topic continuation, are shown in Fig. 11.

The leftmost panel in Fig. 11 shows that when responding to questions about unambiguous full NPs, children do not perform adult-like yet even at age 8. The middle panel in Fig. 11 shows that children do not have a clear antecedent preference for overt pronouns: 6 to 8-year-old children show 45% to 57% topic continuations (average 52%). To test whether this overall mean differed from chance level, a Bayes Factor analysis was done comparing the evidence that this mean is equal to 50% (null hypothesis) to the
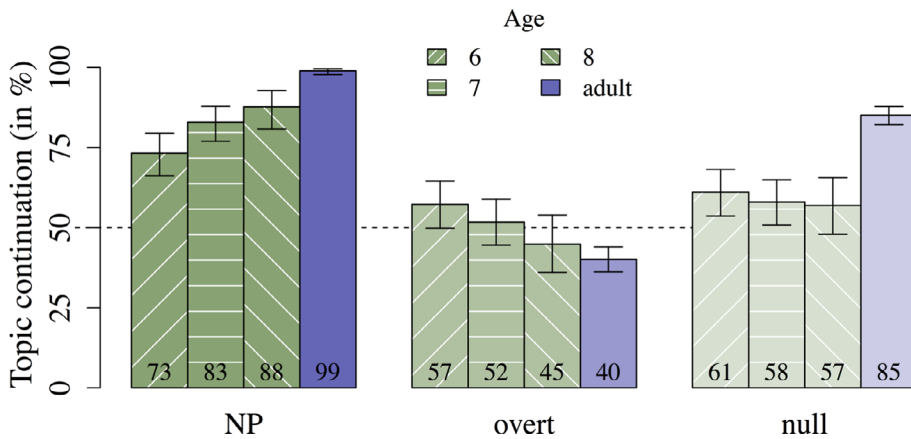
Fig. 11. Experimental data of our child participants and the adult participants of Vogelzang et al. (2020), showing the percentage of responses indicating a topic continuation for the three types of anaphoric subjects in Italian (NP: full NP; overt: overt pronoun; null: null pronoun) per age group (6-year-olds, 7-year-olds, 8-year-olds, and adults). Error bars are derived from logistic analyses.

evidence that the mean is not 50% (alternative hypothesis, cf. Rouder, Speckman, Sun, Morey, & Iverson, 2009). The results show that the evidence for the null hypothesis is greater than the evidence for the alternative hypothesis, and thus that there is evidence that children's interpretations of overt pronouns do not differ from chance level ($BF_{10} = 0.18$). Thus, the children may be guessing when interpreting overt pronouns. The rightmost panel in Fig. 11 shows that 6- to 8-year-old children respond to null pronouns with 57% to 61% topic continuations (average 59%). Again, a Bayes Factor analysis was done comparing the evidence that this overall mean is equal to 50% (null hypothesis) to the evidence that the mean is not 50% (alternative hypothesis). The results show that the evidence for the alternative hypothesis is greater than the evidence for the null hypothesis, so there is evidence that children's interpretations of null pronouns are different from chance level ($BF_{10} = 7.31$). Thus, children preferably interpret null pronouns as referring to the discourse topic.

The observation that the children may be guessing when interpreting overt pronouns is supported by the individual children's responses. Fig. 12 shows that many children interpret overt pronouns about half of the time as referring to the topical referent and about half of the time as referring to the non-topical referent.

The main analysis examines whether children interpret anaphoric subjects in an adult-like manner. For this, we combined the child and adult datasets, and distinguished between the age groups *6-year-olds*, *7-year-olds*, *8-year-olds*, and *adults*. The responses were analyzed using a binomial generalized linear mixed effect-based model. Based on model comparisons, subject form and age group were included as fixed effects, as well as an interaction between subject form and age group. The best model included random intercepts for participants and for items. Random slopes did not
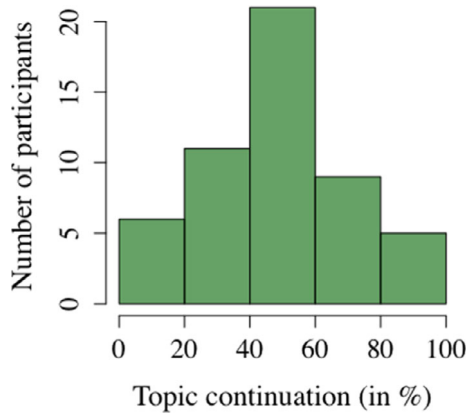
Fig. 12. The distribution of children per mean percentage of topic continuation responses for overt pronouns.

improve the model. In order to compare all different conditions and age groups to each other, we ran multiple models; the cutoff for significance was therefore adjusted to $p < 0.017$.

The model results show that in the baseline condition of an unambiguous NP, all child groups gave fewer correct, topical responses than adults (adults vs. 6-year-olds $\beta = -3.68$; $z = -7.71$; $p < 0.001$; adults vs. 7-year-olds $\beta = -2.99$; $z = -6.18$; $p < 0.001$; adults vs. 8-year-olds $\beta = -2.68$; $z = -5.05$; $p < 0.001$). Furthermore, 6-year-olds gave fewer correct responses than 8-year-olds to questions about full NPs ($\beta = 1.0$; $z = 2.56$; $p < 0.017$). No significant differences were found between the responses of 6-year-olds and 7-year-olds ($\beta = 0.68$; $z = 2.09$; $p = 0.04$), and 7-year-olds and 8-year-olds ($\beta = 0.32$; $z = 0.79$; $p = 0.43$). These results indicate a learning effect over age, with older children responding more adult-like to questions about unambiguous NPs.

When looking at overt pronouns, the analyses show that 6- and 7-year-old children interpreted overt pronouns more often as referring to the topic than adults (resp. $\beta = 0.83$; $z = 3.36$; $p < 0.001$; and $\beta = 0.59$; $z = 2.43$; $p < 0.017$). No difference between 8-year-old children and adults was found ($\beta = 0.26$; $z = 0.91$; $p = 0.36$). Thus, the interpretation of overt pronouns becomes more adult-like with age, and at age 8 children's interpretation is not significantly different from that of adults. Younger children interpret overt pronouns more often as referring to the topic than adults, who prefer a non-topical interpretation of overt pronouns.

Regarding null pronouns, our analyses show that all child groups interpreted null pronouns as referring to the topic less often than adults (adults vs. 6-year-olds $\beta = -1.47$; $z = -5.63$; $p < 0.001$; adults vs. 7-year-olds $\beta = -1.60$; $z = -6.22$; $p < 0.001$; adults vs. 8-year-olds $\beta = -1.68$; $z = -5.70$; $p < 0.001$), but no differences were found between the different child groups (all $ps > 0.017$). This indicates that children's interpretation of null pronouns is not yet adult-like at age 8.

4.3. Discussion

Our results show that Italian 6- to 8-year-olds are still in the process of learning to interpret pronouns. Even though 8-year-olds' performance cannot be distinguished from adults' when interpreting overt pronouns, they are not yet adult-like in their interpretation of full NPs and null pronouns.

Two hypotheses based on our cognitive model were tested in this experiment. The first hypothesis was that children have difficulty keeping track of discourse referents due to limited WM capacity. We predicted that this would lead to mistakes in retrieving the referents of unambiguous full NPs as well as to mistakes in identifying the discourse topic, resulting in fewer topical interpretations for null pronouns compared to adults. These predictions were supported by the data in two ways. First, children showed less accurate performance than adults when interpreting unambiguous NPs, but improve with age. Thus, processing that appears to be simple, such as the processing of full NPs, may still be error-prone in children. This is due to factors that are unrelated to pronoun processing but could be inherent in sentence processing in general, such as attention and WM capacity. Second, children interpreted null pronouns as referring to the discourse topic less often than adults. Contrary to our predictions, children showed no increase in topical, adult-like interpretations of null pronouns. One explanation for this finding is that the presence of a null pronoun has to be derived from the sentence context, which may still be difficult for children (see Section 7 for further discussion).

The second hypothesis was that children have difficulty taking the perspective of the speaker when interpreting overt pronouns. The experimental results are in line with this prediction and show that children interpret overt pronouns more often as referring to the topic than adults, but that they become more adult-like with age. This does not mean that children are not aware of the possibility that referring expressions can refer to non-topical referents. For example, Vernice and Guasti (2014) found that children from the age of 5 can refer both to the topical and a non-topical referent in a sentence continuation task, depending on the prosodic cues that were offered. The findings of Guasti and Chierchia (1999/2000) show that children from the age of 4 show both anaphoric and exophoric interpretations of null pronouns, depending on the context. Overt pronouns may refer to the discourse topic if they are stressed. However, it is unlikely that children's interpretations of overt pronouns were influenced by an incomplete acquisition of the distinction between stressed and unstressed pronouns, since children are able to use contrastive stress during online comprehension in discourse from the age of 5, at least in English (Lee & Snedeker, 2016). It is conceivable, however, that the repetition of the topical full NP influenced the overall coherence of the stories in the experiment and thus influenced the interpretation of overt pronouns, although no evidence of a so-called repeated name penalty (see Gordon, Grosz, & Gilliom, 1993) was found in adults (Vogelzang et al., 2016, 2020).

In the next section, we will present cognitive simulations of Italian pronoun processing in children. These simulations will test various hypotheses regarding children's development of pronoun processing, which will be described below.

5. Modeling study 2: Children's processing of pronouns

Using the same cognitive model as presented in Section 3, but with different parameter settings and amounts of training data to account for differences in WM capacity and processing speed, we run child simulations that will be compared to children's experimental data. In doing so, we adopt the strongest position possible and assume that children's non-adult-like pronoun interpretation is caused by general cognitive limitations, and not by immature linguistic knowledge or immature linguistic processing.

Simulations with the same materials as used for the adult simulations are run in three age groups (6-, 7-, and 8-year-olds), in order to investigate the developmental patterns that were found in the experimental data. We test two main hypotheses. First, we hypothesize that the increase in children's performance on full NPs in our experimental data (described in Section 4) is due to WM capacity increasing with age. Second, we predict that the increase in children's performance on overt pronouns with age can be explained by increased language experience and therefore increased processing speed.

5.1. Developing WM capacity

As discussed in Section 3, WM capacity is implemented in the model as spreading activation from the grammatical subject to the discourse referents associated with it (cf. Daily et al., 2001; van Rij et al., 2013). When WM capacity in the model is limited, as we hypothesize to be the case for children, the grammatical subject will spread less activation, causing the two discourse referents to have a more similar activation level. Thus, without spreading activation, there is no referent that is clearly the most prominent. Consequently, the model will have difficulty determining the discourse topic on the basis of activation. To simulate children's developing WM capacity, the ACT-R mechanism of spreading activation is gradually increased, with no spreading activation at age 6 (cf. Van Rij et al., 2013) to a fifth of adults' amount of spreading activation at age 8 (see Table A1 in the Appendix for the parameter settings per simulated age group).

Fig. 13 shows a plot of the activation of the two discourse referents in memory for the simulated 6-year-old children when the pronoun in the final clause is interpreted as referring to *the hedgehog*. Note the difference between Fig. 2 (adults) and Fig. 13 (6-year-old children): For the simulated adults in Fig. 2, *the hedgehog* clearly has a higher activation than *the mouse* and hence can be identified as the topic, whereas for the simulated 6-year-old children the activation of the two discourse referents is more similar. As a consequence of this similarity in activation, combined with the noise added to retrievals by the ACT-R framework, these simulated children will make mistakes when determining the discourse topic.

Fig. 14 shows the activation of the two discourse referents in memory for the simulated 6-year-old children in a story in which the pronoun in the final clause is interpreted as referring to the other referent, *the mouse*: It can be seen that the activation of *the mouse* (red line) rises only slightly after encountering the pronoun. The activation plots
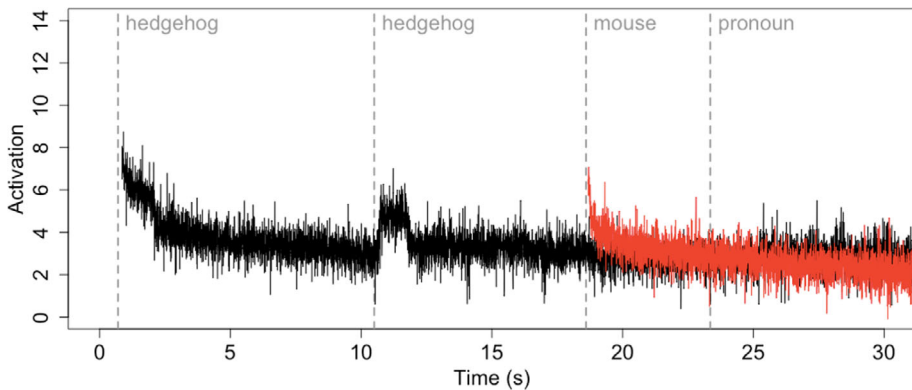
Fig. 13. Referent activation values in memory over time in a low WM capacity model, simulating a 6-year-old child. The black line shows the activation of the first introduced referent (here, *the hedgehog*); the red line shows the activation of the other referent (here, *the mouse*). Because the first introduced referent becomes the highest activated referent after the pronoun is encountered, this simulation illustrates topic continuation.
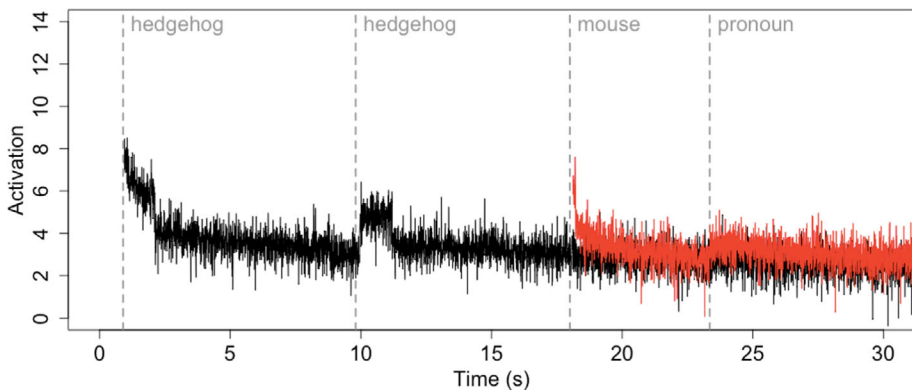

Fig. 14. Referent activation values in memory over time in a low WM capacity model, simulating a 6-year-old child. The black line shows the activation of the first introduced referent (here, *the hedgehog*); the red line shows the activation of the other referent (here, *the mouse*). Because the first introduced referent is not the highest activated referent anymore after encountering the pronoun, this simulation illustrates a topic shift.

in Figs. 13 and 14 show that for simulated 6-year-olds with low WM capacity, it will be difficult to identify the discourse topic.

5.2. Developing processing speed

Executing production rules and retrieving information from memory efficiently is essential in order to fully complete all steps involved in pronoun processing before time

runs out. If the model does not have sufficient time, not all pronoun processing steps can be completed. In that case, the model will select the optimal meaning at that point in time as the output meaning. If no such meaning is available, an output meaning will be selected randomly.

Based on the ACT-R mechanisms of production compilation and the retrieval time associated with chunk activation (see Section 3), the model's processing speed increases through language experience. Experience in the model is provided by training items: Children are simulated by providing the child model with fewer training items than the adult model. The older the simulated child, the more training items it will have received before the test phase. Thus, children of various ages are simulated (see Table A1 in the Appendix for the amount of training items per simulated age group).

5.3. Procedure

The model was run for 20 simulations per age group, simulating 20 children per group (6-, 7-, and 8-year-olds). In each simulation, the model was trained on overt pronouns and null pronouns first, without receiving feedback on the correctness of the responses. Following the training phase, 48 items were presented in the test phase.

5.4. Results

The results of the simulations of the child model are shown in Fig. 15, together with the experimental data on Italian children discussed in Section 4 (model fit Pearson $r^2 = .88$). In line with the experimental data, the simulations show an increase in the correct interpretation of full NPs with increasing age (leftmost panel). The simulations also show that older children have a more adult-like topic shifting interpretation of overt pronouns (middle panel). Furthermore, children's interpretational preference for null pronouns as referring to the topic is replicated by the model (rightmost panel). Unlike the experimental data, however, the model shows an increase in topic continuation interpretations for null pronouns.

5.5. Discussion

Our first hypothesis was that WM capacity, increasing with age, accounts for the increase in children's performance on unambiguous full NPs. The results of the model support this hypothesis and suggest that WM capacity is needed for retrieval from memory. Unfortunately, no WM task was administered in our experimental paradigm, which is a limitation of the current study. The second hypothesis was that language experience, increasing with age and resulting in faster processing speed in the model, accounts for the increase in children's adult-like interpretations of overt pronouns found in the experimental data. The results of the model also support this hypothesis and suggest that increasing processing speed can lead to an increasing ability to perform perspective taking, which was hypothesized to be needed for the mature interpretation of overt pronouns.
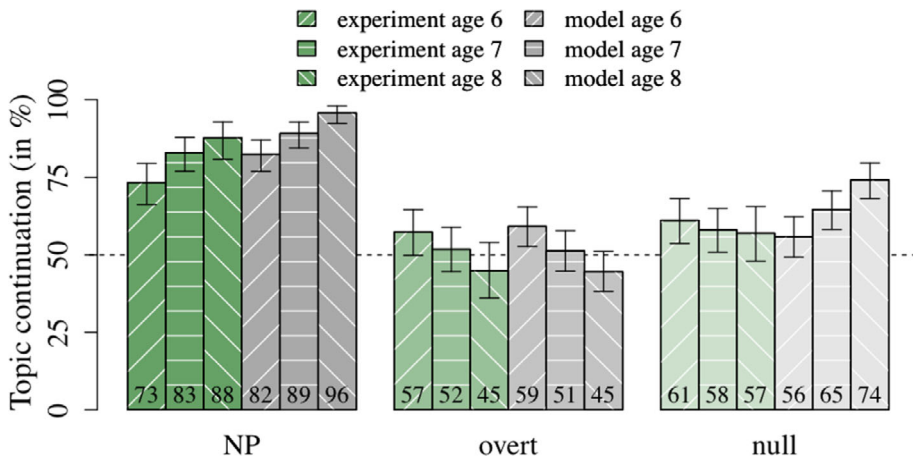
Fig. 15. Experimental data of child participants and corresponding simulation data for three age groups (6-, 7-, and 8-year-olds), for the interpretation of three types of anaphoric subjects in Italian (NP: full NP; overt: overt pronoun; null: null pronoun). Error bars are derived from logistic analyses.

Finally, the model captures the interpretational preference that null pronouns refer to the discourse topic, but also shows an increase in topical interpretations for null pronouns with age that was not found in the experimental data. Null pronouns thus appear to be difficult for developing children to process despite their increasing language experience and WM capacity. Possible explanations of this unexpected pattern will be discussed in Section 7.

Overall, our simulations show that the cognitive model can account for Italian children's interpretational preferences for overt and null pronouns. The simulations match their developmental patterns for NPs and overt pronouns, but not for null pronouns.

6. Novel model predictions

An important aspect of cognitive models is that they can generate novel predictions, which can be tested empirically in future experiments. Based on the adult and child simulations, several predictions can be formulated relating to the influence of WM capacity and processing speed on pronoun processing.

First, the model predicts that when processing speed is low, as in children, but the model receives enough time for processing, performance on overt pronoun interpretation improves. Thus, if the story would be presented to children at a slower speech rate and hence they would have more time to process the pronoun, they would be more likely to complete perspective taking. Indirect evidence for this hypothesis can be found in the study of Van Rij et al. (2010). For a related linguistic phenomenon, namely Dutch children's interpretation of object pronouns in a syntactic binding environment, Van Rij et al.

found that the amount of perspective-taking responses increased with slower speech. Importantly, slower speech is predicted to mainly improve the interpretation of overt pronouns in Italian, and not null pronouns, as we have argued that the time-consuming process of perspective taking is necessary for the interpretation of overt pronouns only.

A second prediction is that when speech rate is higher than usual, adults' interpretations of overt pronouns will become more child-like again. In contrast, a high speech rate should not influence discourse processing much, so adults' performance on null pronouns and full NPs is expected to remain unchanged.

Regarding discourse processing, our model predicts that children who have low WM capacity (e.g., 6-year-olds) will be more affected by the recency and frequency of referents than adults, who have a higher WM capacity that effectively overrules the effects of recency and frequency through spreading activation from the subject. Therefore, young children will be more likely to interpret a recently mentioned referent as the discourse topic. The effects of recency and frequency on referent prominence in children could be tested experimentally. Moreover, WM capacity can be assessed experimentally, and thus its influence on sensitivity to recency and frequency could be investigated.

A fourth and final prediction of the model follows if WM capacity is limited but processing speed is high, such as when adults are engaged in an additional task such as driving a car (see, e.g., Becic et al., 2010). In such a dual task setting, the additional task reduces the amount of WM capacity available for linguistic processing, while not affecting processing speed. This would mainly affect adults' processing and interpretation of null pronouns, due to mistakes in identifying the discourse topic. Such mistakes are predicted to have a smaller effect on the interpretation of overt pronouns, as their interpretations in normal circumstances are close to 50% topic continuation.

7. General discussion

In this paper, we investigated how reduced forms in Italian are processed and acquired, and how this is influenced by cognitive limitations. We implemented a cognitive model in the cognitive architecture ACT-R with which we simulated adults' and children's processing of anaphoric subjects in Italian. Additionally, we performed an empirical experiment testing children's interpretation of anaphoric subjects in Italian.

With the adult simulations, two hypotheses were tested. First, we hypothesized that adults have little problems processing the linguistic discourse and determining the discourse topic. Second, we hypothesized that the interpretation of overt pronouns is derived from the interpretation of null pronouns through perspective taking. The model output was in line with both hypotheses, showing that (a) null pronouns can be resolved solely based on the discourse, and (b) adults' overt pronoun interpretation can be accounted for by perspective taking. This suggests that the processing of null and overt pronouns is dependent on different cognitive capacities: Discourse processing is influenced by WM capacity, and perspective taking is influenced by processing speed.

On the basis of the adult simulations, hypotheses about children's performance were formulated. We first hypothesized that children need sufficient WM capacity to keep track of all discourse referents. Second, we hypothesized that they need sufficient language experience to gain sufficient processing speed to complete the perspective-taking step required for the processing of overt pronouns. Based on these hypotheses, we predicted that increased WM capacity, gained through development, would lead to more adult-like performance on full NPs and null pronouns, and that increased processing speed, gained through language experience, would lead to more adult-like performance on overt pronouns.

The results of both the experiment with Italian children and the child simulations showed, in line with the predictions, that Italian children become more adult-like in their interpretation of full NPs with age. Also as predicted, adult-like responses to overt pronouns, which require the listener to consider the production perspective, increased with age. An unexpected experimental finding was that children did not become more adult-like in their interpretation of null pronouns, which do not require the listener to consider the production perspective. The child model did not capture this experimental finding. It thus seems that children between age 6 and 8 do not benefit from their developing cognitive capacities when processing and interpreting null pronouns. This is a puzzling result, for which we discuss three possible explanations.

A first possible explanation is that adults use predictive parsing (as suggested in computational linguistics by, e.g., Demberg, Keller, & Koller, 2013), with which they predict the syntactic structure of the upcoming sentence. Potentially, adults could use this to predict an upcoming subject after hearing the temporal conjunction *mentre* "while." If children do not yet make these predictions about the syntactic structure of the sentence, they may not be able to predict a null subject before encountering the finite verb. If they only notice the presence of a null subject after having processed the verb, this would leave them with too little time to process the null subject before the next word in the sentence is encountered.

A second, related, possible explanation for the experimental findings on null pronouns is that null pronouns are not always fully recognized when encountered. However, this explanation seems at odds with the finding that, in their own productions, children show evidence of the mastery of null subjects already early on (Guasti & Chierchia, 1999/2000; Vernice & Guasti, 2014), and pronominal features are encoded in the verb morphology, which is also acquired at an early age (Pizzuto & Caselli, 1992).

A third possible explanation is that the specific discourse used in the experiment aided adults' but not children's interpretation of null pronouns. Children are known to have difficulty using discourse cues like the first-mention bias when interpreting pronouns (Arnold et al., 2007). In our experiment, the most prominent referent was the subject of the preceding two clauses, was the first introduced referent, and it was the most frequent referent. The non-topical referent, however, was the most recent referent when the pronoun is encountered. If for children, unlike adults, recency is more important than the other discourse cues, this may have influenced their selection of the discourse topic. Serratrice (2007), using materials in which the discourse topic was less prominent in terms

of subjecthood and frequency than in our experiment, found that Italian children and adults interpret null pronouns as referring to the discourse topic only around 50% of the time. This suggests that differences in discourse prominence could result in differences in null pronoun interpretation. Thus, the idea that children determine discourse prominence differently from adults could partly explain why the children in our study are not adult-like in their interpretation of null pronouns. This idea is consistent with a study on children's use of referring expressions reported in Torregrossa, Bongartz, and Tsimpli (2019), which argues that for children discourse factors (such as distance and number of intervening referents) have more impact than grammatical factors (such as grammatical function). The finding that the 6- to 8-year-old children in our study do not improve in their interpretation of null pronouns with age suggests that the relative importance of the cues for discourse prominence is acquired late.

Irrespective of the cause of children's difficulty with null pronoun interpretation, they will eventually achieve mature performance. Therefore, a next step would be to examine at what age Italian children arrive at a mature interpretation of null pronouns, and whether this is a gradual or a sudden development.

The model's pronoun processing abilities are driven strongly by the process of perspective taking. This process, which accounts for the interpretation of Italian overt subject pronouns in our model, is in line with the widely accepted view in linguistics that, cross-linguistically, the most economical referring expression (i.e., the form with the least phonetic and semantic content) in a language tends to refer to the discourse referent that is most prominent (i.e., accessible, see Ariel, 1990; topical, see Givón, 1983; or given, see Gundel et al., 1993), and less economical forms refer to less prominent referents. In English, the most economical form that can be used in subject position is an overt pronoun; indeed, English overt pronouns tend to refer to the discourse topic. In contrast, in a null-subject language such as Italian, the most economical form in subject position is a null pronoun. Indeed, Italian null pronouns tend to refer to the discourse topic, and Italian overt pronouns generally refer to a less prominent referent, which we argue to be the result of perspective taking.

As mentioned above, our modeling results and experimental results on children's development of pronoun interpretation in Italian support an account in terms of perspective taking. Alternative explanations in terms of ambiguity, markedness, or topic shift avoidance seem less appealing. If children's difficulty with unstressed overt pronouns in Italian would be caused by their ambiguity, it remains unexplained why null pronouns, which in principle are even more ambiguous due to their complete lack of phonetic and semantic content, are less difficult to interpret and less costly to process than overt pronouns (see also Vogelzang et al., 2020). In our perspective-taking account, markedness does play a role, but not as a property of an individual form but rather as a feature of the emerging linguistic pattern in the language: The pattern in Italian shows Horn's division of pragmatic labor (Horn, 1984), in the sense that unmarked forms in the language are used to convey unmarked meanings and marked forms are used to convey marked meanings. Finally, an explanation of children's difficulty with Italian overt pronouns in terms of topic shift avoidance seems unlikely, too, because that would imply that children

generally prefer a topic continuation interpretation and do not distinguish between null and overt pronouns. However, we found that, whereas the Italian children had a preference for reference to the discourse topic for null pronouns, their interpretations of overt pronouns did not differ from chance. This difference in interpretation between null and overt pronouns follows from the perspective-taking account.

The process of perspective taking in sentence processing was first implemented in ACT-R in a model accounting for children's interpretations of object pronouns (see Hendriks & Spenader, 2006, for an OT account of object pronouns, and Hendriks et al., 2007 and Van Rij et al., 2010, for cognitive models based on this account), and it was later extended to subject pronouns (see Hendriks et al., 2008, for an OT account of subject pronouns, and Van Rij et al., 2013, for a cognitive model of subject pronoun processing based on this account). This extension demonstrates that the core of the model is task-independent, with the same processes explaining two apparently different linguistic phenomena, namely the syntactic phenomenon of object pronoun binding and the pragmatic phenomenon of subject pronoun use. Thus, the proposed approach can be generalized to other linguistic phenomena than the one it was originally designed for, and it may therefore also be successfully used in other languages and other linguistic domains in which alternative forms or meanings compete.

In the same way that the model at its core is task-independent, the linguistic account on which the model is based is framework-independent and could potentially be implemented in other computational frameworks than ACT-R. With this in mind, we can compare our model and linguistic account of pronoun processing to other computational models of referential communication found in the literature. The model presented here shows similarities to the work of Kehler and Rohde (2013, 2019), who argue that pronoun interpretation is subject to semantically and pragmatically driven contextual biases and use Bayesian principles to model this. Because in their model pronoun production is insensitive to these contextual biases, their model explains differences between pronoun interpretation and pronoun production. Such differences can also be explained by the linguistic account used in our model (see, e.g., Hendriks & Spenader, 2006), based on whether the relevant linguistic constraints apply to forms or meanings or both. However, the models of Kehler and Rohde do not take the cognitive abilities and limitations of speakers and listeners into account. In addition to assuming a role for contextual information, our model also assumes that listeners take into account the speaker's perspective. This view is in line with the Rational Speech Act framework of Frank and Goodman (RSA; Frank & Goodman, 2012; Goodman & Frank, 2016), in which listeners perform pragmatic reasoning about the speaker's intended meaning in a particular context, based on the utterance used by the speaker and the common knowledge that listeners and speakers share. Like the models of Kehler and Rohde, but unlike our hybrid model that combines symbolic rules with subsymbolic processing, these RSA models are entirely probabilistic and use Bayesian principles. Overall, these alternative models, like our model, place a strong emphasis on the informativeness of words and sentences in context, but they are not concerned with cognitive constraints on referential communication. Thus, whereas these alternative models present *rational* approaches to referential

communication, our model presents a *bounded-rational* approach to referential communication. Further research is needed to investigate how well these alternative models can account for the acquisition of reference in language.

In this study, we used cognitive modeling to investigate the processing of reduced forms such as overt and null pronouns. The outcomes of our model show that the implemented principles are sufficient to explain the observed behavior. However, cognitive models are an abstraction of reality, and therefore a number of assumptions and simplifications were made.

First, not all discourse factors influencing pronoun processing were modeled, but only recency, frequency, and subjecthood. A more complete model of pronoun processing would also take additional discourse factors into account, such as thematic roles (e.g., Arnold, 1998).

Second, the stories were offered to the model in a slower pace (approximately 0.9 words per second) than to the human participants in the experiments (with an estimated speech rate of 1.6 words per second), giving the model slightly more time to process a word. We emphasize that timing in the model indicates relative rather than absolute time. A timing more similar to that in the experiment can be achieved by providing the model with more experience, as that would speed up internal processing. When modeling adults, we used 2,000 training items per simulated adult, which is arguably low compared to the number of pronouns people typically encounter throughout life. Because the model took very long to run, we opted for decreasing the number of training items and increasing the time limit. Nevertheless, the absolute timing of language processing in ACT-R is a topic for debate, as linguistic processes are often faster than the cognitive assumptions of the ACT-R architecture allow (similar issues were mentioned in Van Rij et al., 2010). In that sense, modeling linguistic processes in ACT-R not only aids our understanding of linguistic processes, but it may also help to shape our understanding of the cognitive capacities required for language processing (see also Vogelzang et al., 2017).

Finally, we kept the processes needed for pronoun resolution the same for adults and children. As a consequence, the children in our study, like adults, are assumed to possess the ability of perspective taking, although they will initially have difficulty completing this process within the limited time available for sentence processing, due to their cognitive limitations. Although this may seem like a simplification, 5- and 6-year-old children are already able to consider their conversational partner's perspective in reference resolution (e.g., Nadig & Sedivy, 2002). In addition, the model contained the same underlying linguistic knowledge (implemented as constraints) for adults and children. It is important to note that we do not assume children to be born with this linguistic knowledge, and the model does not provide proof that adults and children possess the same linguistic knowledge. Rather, the model simulates the transition from the point in children's development at which they are not yet able to interpret pronouns in an adult-like way but are already able to correctly interpret full NPs, to the point in their development at which they show adult-like performance in pronoun interpretation. It is known that children have a lower WM capacity than adults (e.g., Gathercole et al., 2004; Van Rijn et al., 2003) and that children have had less linguistic experience than adults (as a result of linguistic

experience being cumulative). This approach thus allowed us to investigate whether the developmental changes in children's pronoun interpretation could be explained by known developments in general cognition, without making additional assumptions about linguistic development. In doing so, we show that the same linguistic knowledge and processes could give rise to child performance and adult performance, as well as the transition between the two, based solely on the interaction between developing WM capacity and increasing linguistic experience. Although the implementation of general principles of cognition in the model and the variation in the cognitive capacities between adults and children of different age groups (i.e., the parameter settings, see Table A1 in the Appendix) were based on previous modeling studies (Daily et al., 2001; Van Rij et al., 2013), more experimental and modeling research is needed to chart the developmental trajectory of these capacities over time, and how this can be most accurately reflected in a cognitive modeling framework.

In conclusion, the cognitive model presented in this paper accounts for the interpretation of overt pronouns, null pronouns, and full NPs in Italian. The cognitive model not only accounts for the general patterns of pronoun interpretation in this language, but also for the differences between adults and children, for the effects of language experience and WM capacity, and for the interplay between these factors. Developmental changes in discourse processing as well as pronoun processing are argued to be explained by children's developing WM capacity and processing speed, with overt and null pronoun processing being influenced by different cognitive factors. Overall, this work demonstrates that the processing of reduced forms can be modeled as an intricate interplay between discourse, linguistic, and cognitive factors, which gives rise to the observed division of pragmatic labor between overt and null pronouns.

Acknowledgments

Notes

1. The model code can be found at https://osf.io/qk54t/?view_only=82509f 428985469c8c859b73e5e2b512.
2. We used the original full set of 120 stories developed by Vogelzang et al. (2020; in the experiment with adults they used 96 of these). We distributed these 120

stories over lists, so that each child was presented with a subset of 60 stories in our experiment. In Vogelzang et al.'s (2020) study, half of the stories were critical items with questions pertaining to the grammatical subject, and the other half of the stories were presented with questions pertaining to the grammatical object. As we replicated their design, each child saw 30 critical stories and 30 stories with questions pertaining to the grammatical object. These stories with questions pertaining to the object can be considered filler items, as only the questions pertaining to the interpretation of the subject are of interest for our investigation. The complete list of stimuli can be found at https://osf.io/qk54t/?view_only=82509f428985469c8c 859b73e5e2b512.

References

Almor, A., Kempler, D., MacDonald, M. C., Andersen, E. S., & Tyler, L. K. (1999). Why do Alzheimer patients have difficulty with pronouns? Working memory, semantics, and reference in comprehension and production in Alzheimer's disease. *Brain and Language*, *67*(3), 202–227. https://doi.org/10.1006/brln.1999. 2055

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?*. New York: Oxford University Press.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, *30*(3), 221–256.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Arnold, J. E. (1998). Reference form and discourse patterns. Doctoral dissertation, Stanford University, Stanford.

Arnold, J. E., 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

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76*(1), B13–B26. https://doi.org/10.1016/S0010-0277(00)00073-1

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748–769. https://doi.org/10.1037//0278-7393.28.4.748

Beaver, D. (2004). The optimization of discourse anaphora. *Linguistics and Philosophy*, *27*(1), 3–56. https:// doi.org/10.1023/B:LING.0000010796.76522.7a

Becic, E., Dell, G. S., Bock, K., Garnsey, S. M., Kubose, T., & Kramer, A. F. (2010). Driving impairs talking. *Psychonomic Bulletin and Review*, *17*(1), 15–21. https://doi.org/10.3758/PBR.17.1.15

Blutner, R. (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics*, *17* (3), 189–216. https://doi.org/10.1093/jos/17.3.189

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2010). The problem state: A cognitive bottleneck in multitasking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 363–382. https://doi.org/10.1037/a0018106

Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue & Discourse*, *2*, 11–33. https://doi.org/10.5087/dad.2011.102

Carminati, M. N. (2002). *The processing of Italian subject pronouns*. Unpublished doctoral dissertation, UMass Amherst.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science*, *25*(3), 315–353. https://doi.org/10.1016/S0364-0213(01)00039-8

De Hoop, H., & Kramer, I. (2006). Children's optimal interpretations of indefinite subjects and objects. *Language Acquisition*, *13*(2), 103–123. https://doi.org/10.1207/s15327817la1302_4

Demberg, V., Keller, F., & Koller, A. (2013). Incremental, predictive parsing with psycholinguistically motivated tree-adjoining grammar. *Computational Linguistics*, *39*(4), 1025–1066. https://doi.org/10.1162/COLI_a_00160

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998. https://doi.org/10.1126/science.1218633

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Givón, T. (1983). *Topic continuity in discourse. A quantitative cross-language study*. Amsterdam: John Benjamins.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*(3), 311–347. https://doi.org/10.1207/s15516709cog1703_1

Gordon, P. C., & Scearce, K. A. (1995). Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory & Cognition*, *23*(3), 313–323. https://doi.org/10.3758/BF03197233

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, *21*(2), 203–225.

Guasti, M. T., & Chierchia, G. (2000). Backward versus forward anaphora: Reconstruction in child grammar. *Language Acquisition*, *8*(2), 129–170. https://doi.org/10.1207/s15327817la82_2

Gundel, J. K., Bassene, M., Gordon, B., Humnick, L., & Khalfaoui, A. (2010). Testing predictions of the Givenness Hierarchy framework: A crosslinguistic investigation. *Journal of Pragmatics*, *42*(7), 1770–1785.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274–307. https://doi.org/10.2307/416535

Hendriks, P., Englert, C., Wubs, E., & Hoeks, J. (2008). Age differences in adults' use of referring expressions. *Journal of Logic, Language and Information*, *17*(4), 443–466. https://doi.org/10.1007/s10849-008-9065-6

Hendriks, P., Koster, C., & Hoeks, J. (2014). Referential choice across the lifespan: Why children and elderly adults produce ambiguous pronouns. *Language and Cognitive Processes*, *29*(4), 391–407. https://doi.org/10.1080/01690965.2013.766356

Hendriks, P., & Spenader, J. (2006). When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition*, *13*(4), 319–348. https://doi.org/10.1207/s15327817la1304_3

Hendriks, P., Van Rijn, H., & Valkenier, B. (2007). Learning to reason about speakers' alternatives in sentence comprehension: A computational account. *Lingua*, *117*(11), 1879–1896. https://doi.org/10.1016/j.lingua.2006.11.008

Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form and use in context (GURT '84)* (pp. 11–42). Washington, DC: Georgetown University Press.

Kaiser, E. (2011). Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus. *Language and Cognitive Processes*, *26*(10), 1625–1666. https://doi.org/10.1080/01690965.2010.523082

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, *25*(1), 1–44. https://doi.org/10.1093/jos/ffm018

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, *39*(1–2), 1–37. https://doi.org/10.1515/tl-2013-0001

Kehler, A., & Rohde, H. (2019). Prominence and coherence in a Bayesian theory of pronoun interpretation. *Journal of Pragmatics*, *154*, 63–78. https://doi.org/10.1016/j.pragma.2018.04.006

Lee, E. K., & Snedeker, J. (2016). Effects of contrastive accents on children's discourse comprehension. *Psychonomic Bulletin and Review*, *23*(5), 1589–1595. https://doi.org/10.3758/s13423-016-1069-7

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

Lorusso, P., Caprin, C., & Guasti, M. T. (2005). Overt subject distribution in early Italian children. In A. Brugos, M. R. Clark-Cotton, & S. Ha (Eds.), *BUCLD 29 Proceedings Supplement*. Boston, MA.

McDonald, J. L., & MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, *34*(4), 543–566. https://doi.org/10.1006/jmla.1995.1025

Misker, J. M. V., & Anderson, J. R. (2003). Combining optimality theory and a cognitive architecture. In F. Detje, D. Dörner, & H. Schaub (Eds.), *Proceedings of the Fifth International Conference on Cognitive Modeling* (pp. 165–170). Bamberg, Germany: Universitäts-Verlag Bamberg.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336. https://doi.org/10.1111/j.0956-7976.2002.00460.x

Pizzuto, E., & Caselli, M. C. (1992). The acquisition of Italian morphology: Implications for models of language development. *Journal of Child Language*, *19*(3), 491–557. https://doi.org/10.1017/S0305000900011557

Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint Interaction in generative grammar*. Oxford, UK: Blackwell. https://doi.org/10.1002/9780470759400

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Serratrice, L. (2005). The role of discourse pragmatics in the acquisition of subjects in Italian. *Applied Psycholinguistics*, *26*(03), 437–462. https://doi.org/10.1017/S0142716405050241

Serratrice, L. (2007). Cross-linguistic influence in the interpretation of anaphoric and cataphoric pronouns in English-Italian bilingual children. *Bilingualism: Language and Cognition*, *10*(3), 225–238. https://doi.org/10.1017/S1366728907003045

Serratrice, L., Sorace, A., Filiaci, F., & Baldo, M. (2012). Pronominal objects in English – Italian and Spanish – Italian bilingual children. *Applied Psycholinguistics*, *33*(4), 725–751.

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, *22*(3), 339–368. https://doi.org/10.1191/0267658306sr271oa

Stewart, A. J., Holler, J., & Kidd, E. (2007). Shallow processing of ambiguous pronouns: Evidence for delay. *Quarterly Journal of Experimental Psychology*, *60*(12), 1680–1696. https://doi.org/10.1080/17470210601160807

Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say "Broke"? A model of learning the past tense without feedback. *Cognition*, *86*(2), 123–155.

Torregrossa, J., Bongartz, C., & Tsimpli, I. M. (2019). Bilingual reference production: A cognitive-computational account. *Linguistic Approaches to Bilingualism*, *9*(4–5), 569–599. https://doi.org/10.1075/lab.17026.tor

Van Rij, J., Van Rijn, H., & Hendriks, P. (2010). Cognitive architectures and language acquisition: A case study in pronoun comprehension. *Journal of Child Language*, *37*(3), 731–766. https://doi.org/10.1017/S0305000909990560

Van Rij, J., Van Rijn, H., & Hendriks, P. (2013). How WM load influences linguistic processing in adults: A computational model of pronoun interpretation in discourse. *Topics in Cognitive Science*, *5*(3), 564–580. https://doi.org/10.1111/tops.12029

Van Rijn, H., Van Someren, M., & Van der Maas, H. (2003). Modeling developmental transitions on the balance scale task. *Cognitive Science*, *27*(2), 227–257. https://doi.org/10.1207/s15516709cog2702_4

Vernice, M., & Guasti, M. T. (2014). Effects of prosodic cues on topic continuity in child language production. *First Language*, *34*(5), 406–427. https://doi.org/10.1177/0142723714550111

Vogels, J., Krahmer, E., & Maes, A. (2015). How cognitive load influences speakers' choice of referring expressions. *Cognitive Science*, *39*(6), 1396–1418. https://doi.org/10.1111/cogs.12205

Vogelzang, M., Foppolo, F., Guasti, M. T., Van Rijn, H., & Hendriks, P. (2020). Reasoning about alternative forms is costly: Comparing the processing of null and overt pronouns in Italian using pupillary responses. *Discourse Processes*, *57*(2), 158–183. https://doi.org/10.1080/0163853X.2019.1591127

Vogelzang, M., Hendriks, P., & Van Rijn, H. (2016). Pupillary responses reflect ambiguity resolution in pronoun processing. *Language, Cognition and Neuroscience*, *31*(7), 876–885. https://doi.org/10.1080/23273798.2016.1155718

Vogelzang, M., Mills, A. C., Reitter, D., Van Rij, J., Hendriks, P., & Van Rijn, H. (2017). Toward cognitively constrained models of language processing: A review. *Frontiers in Communication*, *2*(11), 1–18. https://doi.org/10.3389/fcomm.2017.00011

Appendix A:

Table A1

Parameter settings for the different amounts of spreading activation (W in formula b) and the different numbers of training items used in the simulation of the four age groups

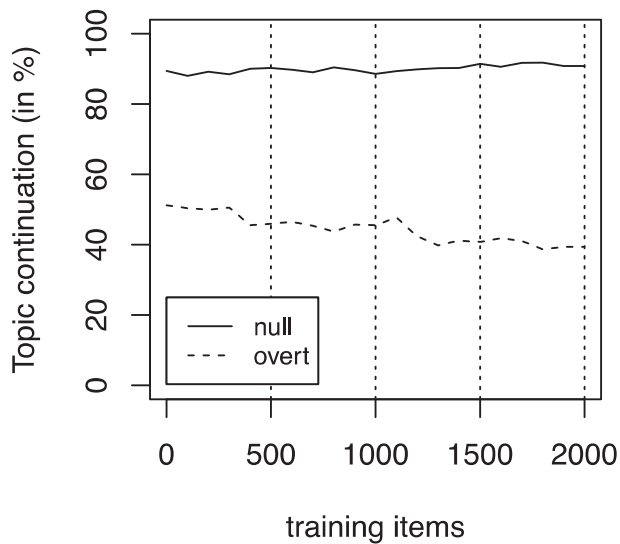| Simulated Age	Spreading Activation (W)	Training Items
6-year-olds	0.0	500
7-year-olds	0.1	1,000
8-year-olds	0.2	1,500
Adults	1.0 (default value)	2,000

Fig. A1. Development in the model of the interpretation of null and overt pronouns as indicating a topic continuation as a function of the number of training items received (means of 40 simulations). Moving average (100 trials) of the percentage of topic continuation outcomes of the pronoun processing part of the model per number of training items. Vertical dotted lines indicate the number of training items presented to the four simulated age groups (6-, 7-, 8-year-olds, and adults).
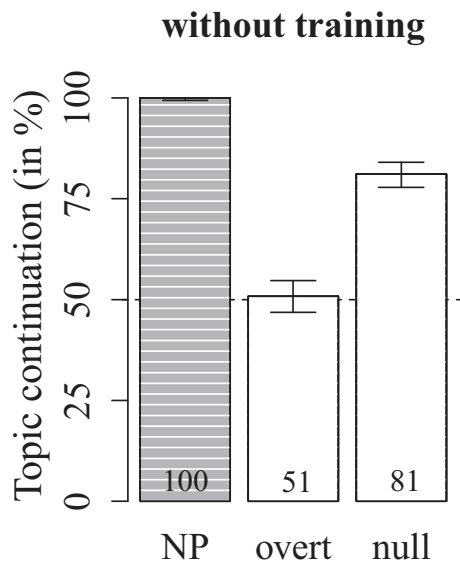

Fig. A2. Simulation data of the adult model on the interpretation of three types of anaphoric subjects in Italian (NP: full NP; overt: overt pronoun; null: null pronoun) with adult WM capacity (i.e., with a spreading activation of 1.0), but without training. Error bars are derived from logistic analyses.
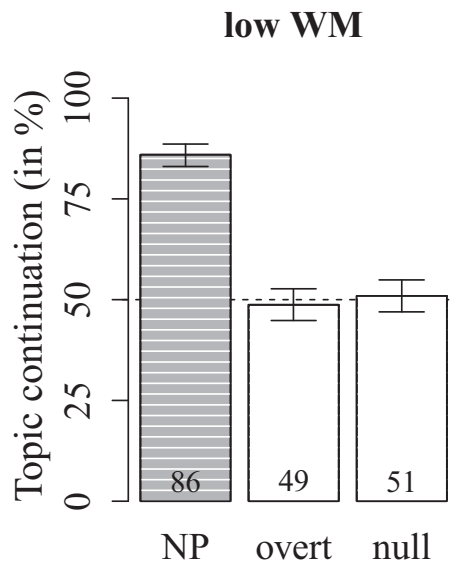
low WM

Fig. A3. Simulation data of the adult model on the interpretation of three types of anaphoric subjects in Italian (NP: full NP; overt: overt pronoun; null: null pronoun) when training has been received (2,000 items), but without adult WM capacity (i.e., without spreading activation). Error bars are derived from logistic analyses.