Pronoun processing

Computational, behavioral, and psychophysiological studies in children and adults

Jacolien van Rij

Groningen Dissertations in Linguistics 108
ISBN printed version: 978-90-367-5920-5
ISBN digital version: 978-90-367-5919-9

The work presented in this thesis has been carried out under the auspices of the Research School of Behavioral and Cognitive Neurosciences (BCN) and the Center for Language and Cognition Groningen (CLCG) of the University of Groningen.

This investigation was partially supported by a grant from the Netherlands Organisation for Scientific Research (NWO) awarded to Petra Hendriks (grant. no. 277-70-005) for the VICI project *Asymmetries in Grammar*.

Cover design by Petra van Berkum, illustrator | www.petravanberkum.nl
Printed by Gildeprint Drukkerijen, Enschede, The Netherlands | www.gildeprint.nl
© 2012, Jacolien van Rij

RIJKSUNIVERSITEIT GRONINGEN

Pronoun processing

Computational, behavioral, and psychophysiological studies in children and adults

Proefschrift

ter verkrijging van het doctoraat in de
Letteren
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
donderdag 29 november 2012
om 12.45 uur

door

Jacomina Cornelia van Rij-Tange

geboren op 26 oktober 1981
te Middelburg

Promotor:	Prof. dr. P. Hendriks
Copromotor:	Dr. D. H. van Rijn
Beoordelingscommissie:	Dr. J. E. Arnold Prof. dr. ir. J. Nerbonne Prof. dr. S. Vasishth

Acknowledgements

Finishing this book is the end of the four-year PhD project. Many people have contributed to this project with ideas, advice, and practical assistance, or indirectly by making my time in Groningen more enjoyable.

First of all, I would like to thank Petra Hendriks and Hedderik van Rijn, who supervised my PhD project and the preceding Master project. Thank you for your enthusiasm, ideas, and support. Petra, your enthusiasm was very inspiring. You always made time for answering my questions, and regularly commented on my work during holidays. Also thanks for encouraging me to visit so many conferences. Hedderik, you came up with many new creative ideas and interesting statistical analyses – I learned a lot from that. Thank you for the possibilities to learn new experimental techniques and methods. I'm also thankful to Jennifer Spenader, who was involved in the early stages of this project. Thank you for helping me to get started in research.

I would like to thank the reading committee for the time they took to read and evaluate this work, and for their valuable comments: Jennifer Arnold, John Nerbonne, and Shravan Vasishth.

The four experiments reported in this thesis could not have been performed without the help and practical assistance of the following people: Petra van Berkum and Robbert Prins for drawing the pictures used in the object pronoun experiments; Margreet Drok and Atty Schouwenaars for recording sentences; Sanne Berends and Sanne Kuijper for assisting with testing children; Charlotte Koster, Gisi Cannizzaro, and Sanne Kuijper for their help in creating all the stories for the subject pronoun experiments; Simone Sprenger for helpful advice on the design of psychophysiological experiments; Tadeusz Kononowicz for teaching me how to do EEG experiments and for explaining the preprocessing of EEG data; Harald Baayen for the useful discussions about statistical analyses of EEG data; Heleen Meijburg, Trudy Buwalda, and Marijke Beulen for assisting with the EEG experiment; and the children, teachers and parents of the Nassauschool in Groningen for participating in the slowed-speech experiment.

During my PhD I had the privilege to participate in two different, but very good research groups. The *Asymmetries in Grammar* NWO Vici project directed by Petra Hendriks provided a friendly working atmosphere. I learned a lot about language acquisition research and eye tracking from discussions and collaborations with Charlotte Koster, Bart Hollebrandse, John Hoeks, Petra Hendriks, Peter de Swart, Gisi Cannizzaro, Sanne Kuijper, Jessica Overweg, Franziska Köder, and the other members. Thank you for the good time.

I also participated in the CogMod Lab, where I learned a lot from discussion about

experimental designs, new experimental methods, and about how to make perfect posters and presentations. Niels Taatgen, Hedderik van Rijn, Jennifer Spenader, Leendert van Maanen, Jelmer Borst, Katja Mehlhorn, Marieke van Vugt, Enkhbold Nyamsuren, Ben Meijering, Tadeusz Kononowicz, Stefan Wierda, Menno Nijboer, Udo Böhm, Trudy Buwalda, and all other members of the group, thank you for your critical questions, feedback, and support.

In addition, I would like to thank the Experimental Linguistic group and the Language Acquisition group for their feedback: Ryan Taylor, Laurie Stowe, Harm Brouwer, Angeliek van Hout, Ruggero Montalto, Susanne Grassmann, Atty Schouwenaars, and other regular visitors of these meetings. Also thanks to Gisi Cannizzaro and Peter de Swart for organizing the Statistical Reading group, which helped me to read difficult, but useful papers and books.

I want to thank a number of other people for their support during my PhD time. My paranimphs, Jelmer and Sanne, for being paranimfs and for your advice on the thesis and the defense day. Jelmer, great that you could schedule your stay in Groningen after the defense date was sure. Sanne, I enjoyed the coffee and tea moments in our/your office. Gisi, thanks for all your *"strenge"* comments on English writings, and for getting me involved in eye tracking research. Veerle, I also appreciated the regular lunch conversations very much. Your practical help and sharing frustrations were invaluable. Katja, working with you at Psychology was always *gezellig*. Many thanks for letting us borrow Thor and Parvati. Martijn, thanks for the interesting conversations about statistics and for all practical things you organized and achieved in the interest of PhD students. Thanks to all people I shared an office with. Especially I would like to thank Leendert and Jelmer, Katja, Veerle and Myrte, Sanne, and Menno and Udo for the good time. Furthermore, thanks to my CLCG colleagues Ankelien and Hanneke for sharing information about practical matters around printing the thesis, and also to Wyke van der Meer, the coordinator of the CLCG, for help with administrative questions.

There is more to life than science. Hennie, Jeroen, Harm Jan, Jorien, Karin, Peter, Tjard, Margreet, Johan, Hieke, Remco, and Emmy, and other friends, thank you for your friendship. You have made our time in Groningen unforgettable. I'm grateful to my parents for encouraging me to go to Groningen, and for their involvement, advice and support all these years. Also I would like to thank my sisters, mother-in-law, and my brothers- and sisters-in-law for their interest and support. Finally, I want to thank Jaco for sharing all these years in Groningen with me. Without your encouragement and support I probably would not have started this project, and without your advice and your ability to put things in the right perspective, I could not have finished it.

Contents

1 Introduction 1

1 Pronoun processing . . . 3
2 The cognitive architecture ACT-R . . . 5
3 Linguistic background . . . 7
4 Overview . . . 10

2 Cognitive architectures and language acquisition: A case study in pronoun comprehension 13

1 Introduction . . . 15
2 Delay of Principle B Effect (DPBE) . . . 16
3 Optimality Theory explanation of the DPBE . . . 19
4 Testing linguistic theories . . . 23
5 Cognitive models of language acquisition . . . 24
6 Study 1: Experimental study . . . 28
7 Study 2: Simulation study . . . 35
8 General discussion . . . 38

3 A cognitive model of the acquisition and use of referring expressions: Effects of working memory and perspective taking 43

1 Introduction . . . 45
2 Empirical phenomena . . . 47
3 A computational model of the production and comprehension of referring expressions . . . 51
4 Results of the simulation . . . 58
5 General discussion . . . 66
6 Conclusion . . . 70

4 How WM load influences linguistic processing in adults: A computational model of pronoun interpretation in discourse 73

1 Interpretation of subject pronouns . . . 75
2 Cognitive model . . . 76
3 Dual-task experiment . . . 79
4 Discussion . . . 84

Appendices . . . 86

5 Context resolves ambiguity before an ambiguous pronoun is read: An ERP study of the effects of context and working memory load on pronoun comprehension 89
1 Introduction . . . 91
2 Experimental materials and methods . . . 95
3 Results . . . 98
4 Discussion . . . 107
Appendices . . . 112

6 Event-related pupil dilation tracks integration of contextual and grammatical information 115
1 Introduction . . . 117
2 Experimental materials and methods . . . 120
3 Results . . . 123
4 Discussion . . . 126
Appendices . . . 129

7 Discussion 133
1 Introduction . . . 135
2 Effects of cognitive factors and linguistic constraints . . . 135
3 Theoretical implications . . . 137
4 Implementation issues . . . 140
5 Conclusions . . . 141

References 143

Nederlandse samenvatting 153

chapter 1

Introduction

1 Pronoun processing

In many contexts, a speaker can choose between several forms to refer to a particular event or individual. However, some referring expressions are more specific than others. Pronouns, such as *he*, *she*, *him*, or *her*, do not have a fixed interpretation. Rather, their interpretation depends on the context, in contrast to the interpretation of proper names such as *Epke Zonderland*. Generally, a pronoun is used when reference is intended to a salient referent in the linguistic context, whereas more specific forms are used when reference is intended to less salient referents or when new referents are introduced. Although most theories of pronoun use would agree with this, they use different notions to capture the referent's prominence in the linguistic context, such as *accessibility* (Ariel, 1988), *givenness* (Gundel, Hedberg, & Zacharski, 1993) and *topicality* (Givón, 1983).

In languages like Dutch and English, pronouns are only specified for case, number and gender information. Therefore, in many situations more than one referent in the context can be the antecedent of the pronoun (i.e., the referent to which the pronoun refers). For example, when encountering a pronoun at the end of a paragraph in a novel with many different characters, you may discover that you are not sure whom the author is referring to. Also in the following sentences the pronoun is ambiguous for most non-Dutch readers, and may refer to Mart Smeets or Epke Zonderland: *Mart Smeets and Epke Zonderland talked about the preparation for the olympic gold winning routing on the acrobatic high bar. He worked very hard to achieve this result.*[1] Interestingly, in most situations people do not experience any difficulties in the interpretation of potentially ambiguous pronouns, but immediately interpret the pronoun the way it was intended by the speaker or writer. In the previous example about the olympic gold medal, all Dutch readers immediately interpret the pronoun as referring to Epke Zonderland.

Pronoun resolution is crucial for sentence comprehension: If you are not able to resolve the pronoun, you may not grasp the meaning of the sentence and hence may misunderstand the entire story or conversation. Therefore, important questions in linguistic and psycholinguistic research are: How do listeners know what is the referent of a pronoun, and what determines a speaker's choice for a pronoun instead of a more informative referring expression, such as a proper name, a definite noun phrase (*the gymnast*) or an indefinite noun phrase (*a gymnast*)? Previous research has shown that several factors play a role in pronoun interpretation and production, such as grammatical principles and the structure of the preceding linguistic context.

Studies on the acquisition of pronouns indicate that cognitive development also plays an important role. Dutch and English-speaking children acquire adult-like pronoun interpretation and production relatively late in language development, generally not before the age of six (a.o., Chien & Wexler, 1990; Spenader, Smits, & Hendriks, 2009; Karmiloff-Smith, 1985; Grimshaw & Rosen, 1990; Philip & Coopmans, 1996; Koster, 1993), suggesting that the adult-like use of pronouns is hard to master. This raises further questions: What makes pronouns so difficult to learn, and what cognitive factors are critical for adult-like pronoun processing? Although pronoun resolution is an important topic in linguistic and psycholinguistic research, there is no general agreement about the underlying mechanisms. It also remains

unclear how the different factors, such as grammatical principles, the linguistic context, and cognitive factors, interact.

1.1 Research questions

This thesis aims to investigate the relative contribution of linguistic constraints and general cognitive factors on the interpretation and production of pronouns. Different cognitive factors have been found to influence language processing, such as working memory capacity, speed of processing, and Theory of Mind-like skills. How do these factors influence pronoun processing? Investigating the interaction between the linguistic constraints and the limitations of the cognitive system that is used for language processing may provide insights in pronoun acquisition and the pronoun processing mechanism.

Different linguistic constraints influence pronoun processing: grammatical principles, such as the principles of Binding Theory (Chomsky, 1981) and discourse factors like the saliency of referents in the linguistic context. Traditionally, the grammatical principles that guide the use of pronouns are studied in the syntactic domain, whereas the effects of the context are studied in the pragmatic domain. In the linguistic framework that is used in this thesis, Optimality Theory (OT; Prince & Smolensky, 1997, 2004), grammatical principles and discourse factors can both be formulated as linguistic constraints. This thesis tries to integrate the insights from syntax and pragmatics by investigating the subquestion of how people combine grammatical principles and discourse information in pronoun processing.

1.2 Research methodology

To investigate pronoun processing, I have implemented a computational model in the cognitive architecture ACT-R (Anderson, 2007; Anderson et al., 2004) that simulates the production and comprehension of referring expressions. The computational model is based on linguistic theories of pronoun production and interpretation (a.o., Hendriks & Spenader, 2006). Implementing a computational model requires the modeler to specify the workings of the theory in detail, and to make all underlying assumptions explicit.

Implementing computational models within a cognitive architecture has additional advantages. A cognitive architecture aims at being a unified theory of cognition that can explain human cognition across a range of different situations. The assumptions of the cognitive architecture are based on the results of numerous psychological and neurophysiological experiments. Implementing a computational model within a cognitive architecture such as ACT-R ensures cognitive plausibility because the implementation is grounded in a general theory of cognition. Additionally, implementing a computational model within a cognitive architecture reduces the number of free parameters because the many simulations that are implemented within the same cognitive architecture together validate the assumptions and parameters of the cognitive architecture. The goal of implementing a cognitive model of pronoun processing is to generate very specific, testable, and cognitively plausible predictions with respect to the production and comprehension of pronouns, based on the model's simulations.

Using cognitive models of pronoun processing embedded in the cognitive architecture ACT-R, this thesis investigates the complex interaction between linguistic constraints and cognitive factors. The predictions of our cognitive model can subsequently be investigated using empirical and psychophysiological experiments. Different off-line and on-line empirical methods can be used to examine these predictions because the assumptions formulated in our cognitive model constrain the interpretation of our results. Note that this methodology does not involve comparing different computational models that implement different theoretical assumptions. Rather, because our predictions are very precise and our assumptions based on the linguistic theories have been made explicit, the results of the empirical studies can be used to validate the cognitive model under investigation, as well as the linguistic theories on which the model is based. In addition, the results may provide new insights in pronoun processing, which could justify adjusting or extending the model. Thus, my methodology is an iterative process that consists of computational modeling, generating testable predictions, testing these predictions with empirical studies, and evaluating the model, as illustrated in Figure 1.

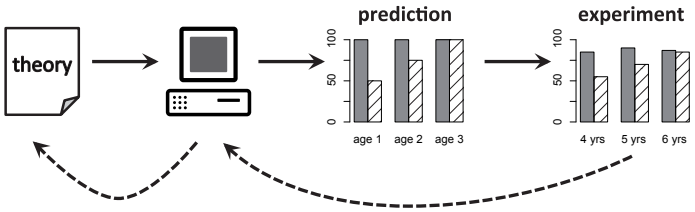


Figure 1: Overview of the methodology used in cognitive modeling.

In the rest of this chapter, I will introduce the cognitive architecture ACT-R in more detail and present the theoretical assumptions that form the basis of this thesis. I will conclude with an overview of the next chapters.

2 The cognitive architecture ACT-R

Cognitive models are computational simulations of the cognitive processes involved in performing a task, such as comprehending a sentence (R. L. Lewis & Vasishth, 2005). The cognitive model presented in this thesis is implemented in ACT-R (Adaptive Control of Thought-Rational; Anderson, 2007; Anderson et al., 2004), which is a cognitive architecture that aims to explain human cognition in different situations and tasks. As such, it is not only a modeling framework, but also a theory of cognition in which a range of assumptions are implemented that are based on more specialized theories of cognition. ACT-R's detailed and well-tested assumptions about the cognitive system make it a suitable framework to investigate the effects of cognitive factors, such as processing speed and working memory capacity, on pronoun processing. ACT-R models may result in cognitively plausible predictions about higher-level cognition, such as behavioral performance and on-line processing in experimental tasks.

2.1 The structure of ACT-R

Figure 2 illustrates the modular structure of ACT-R. The modules are based on smaller theories of cognition. The two modules of ACT-R that are most relevant for this thesis are the declarative memory and the central production system.

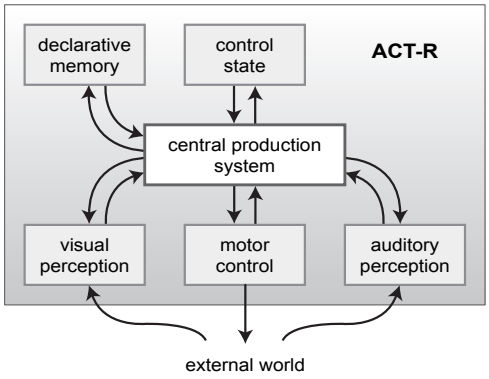


Figure 2: The modular structure of ACT-R (adapted from Anderson et al., 2004).

The declarative memory module contains chunks of factual information, such as the linguistic knowledge necessary for the interpretation and production of pronouns. Chunks can be connected with other chunks, forming a network. Each chunk has an activation value that reflects its own history and its relevance in the current context. The activation of chunks determines both the probability that this chunk will be retrieved and the time required for retrieval. In the ACT-R framework, the central production system contains IF-THEN rules, which implement strategies or procedures that language users apply in sentence processing and production. The production rules issue the retrieval of information from the declarative memory and subsequently manipulate that information for further use. For example, these production rules may implement grammatical principles that guide sentence processing (cf. R. L. Lewis & Vasishth, 2005). Note that the choice for firing a production rule depends on whether the IF-clause of the production rule matches the state of the current task and the usefulness of that production rule in the past, as reflected by the utility of the production rule.

The assumptions of the ACT-R architecture determine the language processing mechanism. A basic assumption of ACT-R is that only one production rule can fire at a time, which requires the language processing models to perform certain actions serially. However, the architecture also assumes that some subprocesses are performed in parallel. For example, while the system is waiting for the retrieval of declarative information, the visual module may focus on new visual information that is coming in. Another example is that the activation of chunks fluctuates continuously depending on their own history and the current context. Thus, ACT-R poses more restrictions on sentence processing with respect to timing and competition between different kinds of information than most constraint-based theories

of sentence processing assume (a.o., MacDonald, Pearlmutter, & Seidenberg, 1994). Although ACT-R poses restrictions on the implementation of sentence processing, it does not enforce a particular strictly modular or serial account of sentence processing, which assumes that first a syntactic analysis of the sentence is constructed, before semantic and discourse information is integrated (a.o., Frazier,1987; see van Gompel & Pickering, 2007, for a review). In Chapter 3, relevant assumptions of ACT-R are explained in more detail. So, by implementing the models in ACT-R, our computational implementation of pronoun processing is constrained by ACT-R's assumptions about cognitive processing.

3 Linguistic background

As the starting point of our implementation of pronoun processing, we use Hendriks and Spenader's (2006) linguistic account of the acquisition of object pronouns. This account provides an explanation of the finding that English-speaking and Dutch-speaking children up to age 6 still make errors interpreting object pronouns (a.o., Chien & Wexler, 1990; Philip & Coopmans, 1996; Koster, 1993). Hendriks and Spenader's Optimality Theoretic account of this comprehension delay in language acquisition predicts that children acquire adult-like production of object pronouns before they acquire adult-like comprehension of object pronouns. This prediction was confirmed by later empirical studies (De Villiers, Cahillane, & Altreuter, 2006; Spenader et al., 2009; Matthews, Lieven, Theakston, & Tomasello, 2009). Hendriks and Spenader proposed that the grammatical principles that guide the production and interpretation of object pronouns are direction-sensitive, causing an ambiguity in the interpretation of object pronouns but at the same time resulting in correct production of object pronouns. Children show this ambiguity in their interpretation of object pronouns and incorrectly allow the object pronoun to corefer with the subject of the sentence as well. Adults, in contrast, use perspective taking to resolve the ambiguity of the pronoun: by taking the perspective of the speaker into consideration they know that if a speaker intended to refer to the local subject, he or she would have used a reflexive. Therefore, adult listeners interpret an object pronoun as reference to a non-subject referent, whereas children also allow the subject as the antecedent of the pronoun.

A similar grammatical account has been proposed for the production of subject pronouns. Here, adult speakers have been argued to consider the listener's perspective when choosing a referring expression to use (Hendriks, Englert, Wubs, & Hoeks, 2008). Support for this account of the use of subject pronouns has been provided by Wubs, Hendriks, Hoeks, and Koster (2009), who found that 4- to 6-year-old children use pronouns even for reference to a not very salient discourse referent. In contrast, adult speakers try to avoid ambiguity for the listener and use a full noun phrase instead.

Using a direction-sensitive grammar, these accounts are able to explain asymmetries between production and interpretation in language acquisition. These asymmetries in language acquisition disappear when children have learned to use perspective taking. We have implemented the theory of Hendriks and Spenader in ACT-R with the aim of explaining why children do not use perspective taking and how exactly they acquire the perspective taking skills

that are necessary for adult-like processing of pronouns.

3.1 The linguistic theory Optimality Theory

Hendriks and Spenader formulated their account within the linguistic framework Optimality Theory (OT; Prince & Smolensky, 2004). In semantics, OT models the relation between an input form and the optimal meaning for that form (e.g., Hendriks & de Hoop, 2001). In syntax, OT models the relation between an input meaning and the optimal form to express that meaning. The OT optimization process for interpretation is illustrated in Figure 3.

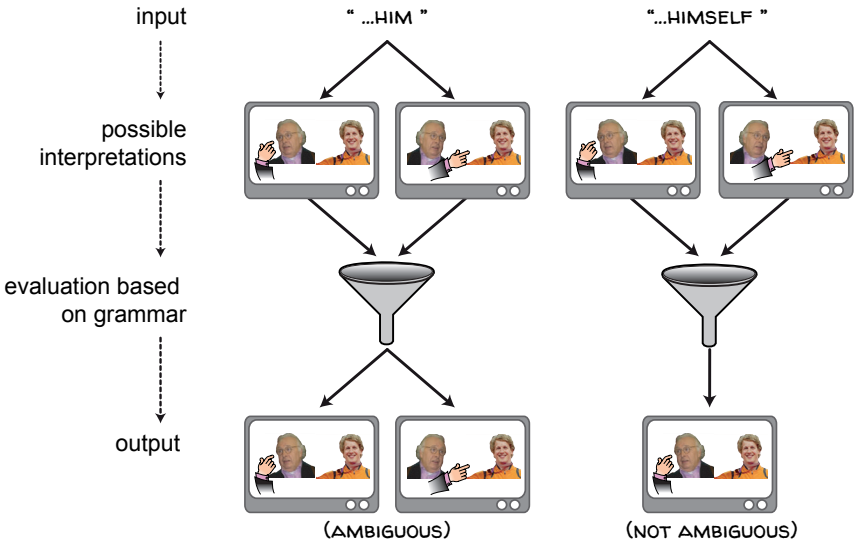


Figure 3: Optimization from input to output in OT (left), illustrated with the interpretation of object pronouns (middle) and reflexives (right) for the sentence *During the interview, Mart Smeets pointed to him / himself*

On the basis of the input, candidate outputs are generated, which are evaluated by the grammar. As illustrated in Figure 3, for pronouns and reflexives several candidate interpretations compete: the pronoun or reflexive may refer to the subject referent (*Mart Smeets*) or to a non-subject referent (*Epke Zonderland*). According to the grammar, reflexives must refer to the subject referent, but pronouns are ambiguous. Pronouns can refer to a non-subject referent as well as to a subject referent. This corresponds to the interpretational pattern shown by Dutch children.

An OT grammar consists of a set of ranked and violable constraints. These OT constraints pertain to syntactic, semantic, and discourse information and may interact. To determine the optimal candidate, these constraints are simultaneously applied to the candidates. The output of the OT grammar is the candidate that commits the least severe constraint viola-

tions. In OT, it is more important to satisfy one highly ranked constraint than several lower ranked constraints. Thus, the optimal output does not necessarily satisfy all constraints. In this respect, violable constraints differ from inviolable rules.

After finding the optimal output(s), the opposite perspective is taken into account to check whether the optimal choice for the listener is also optimal for the speaker and vice versa. In OT this process of perspective taking is modeled as bidirectional optimization (Blutner, 2000). In bidirectional optimization, both the listener's perspective and the speaker's perspective are relevant for determining the optimal form or meaning. Through this process, adults are hypothesized to block reference of an object pronoun to the subject referent.

3.2 Implementation of OT account

This OT account is implemented within the cognitive architecture ACT-R in accordance with the assumptions of ACT-R. For example, as the declarative knowledge in ACT-R is not ranked but rather organized in networks of connected chunks, we have not implemented a strict hierarchical ranking of grammatical constraints. Instead, the ranking of the grammatical constraints is reflected in the activation of the chunks representing the grammatical constraints (but see Misker & Anderson, 2003, for a different implementation of OT in ACT-R). Furthermore, we used ACT-R's learning mechanisms to explain the acquisition of object pronouns because OT does not specify the learning mechanism for acquiring adult-like bidirectional behavior (although learning mechanisms have been proposed in OT to account for children's reranking of constraints, see Tesar & Smolensky, 1998; Boersma & Hayes, 2001).

The cognitive model uses the same grammatical constraints for the interpretation and production of object pronouns, as proposed by Hendriks and Spenader. In addition, the model implements bidirectional optimization as two subsequent processes of unidirectional optimization. In the first process the model takes its own perspective as a listener (interpretation) or speaker (production), and in the second process the model takes the opposite perspective.

3.3 Generalizing the cognitive model

In Chapter 2, I will present a cognitive model that explains the acquisition of object pronouns. In Chapter 3, the model is extended to explain the acquisition of subject pronouns. Traditionally, *subject pronouns* are studied in the domain of pragmatics, which focuses on the effects of the preceding discourse, whereas *object pronouns* are studied in the syntactic domain, which focuses on the grammatical principles that guide sentence interpretation. However, the Optimality Theoretic accounts explaining the acquisition of object pronouns and subject pronouns are very similar because both delays in acquisition are explained from children's difficulty with perspective taking. Based on these accounts, I implemented a cognitive model of subject pronouns using the same underlying mechanism as was used in the cognitive model accounting for the acquisition of object pronouns. In addition, insights from pragmatic theories of pronoun use were implemented, such as the influence of discourse on the saliency of the referents in the linguistic context (a.o., Ariel, 1988; Arnold, 1998; Givón, 1983). The extended

model thus integrates a grammatical account and a discourse account of pronoun use, and may also be used to explain the processing of subject pronouns and object pronouns.

4 Overview

In this thesis, I present a cognitive ACT-R model as a novel account of the interpretation and production of referring expressions in discourse, and discuss the predictions that follow from computational simulations and the results of the empirical studies that are performed to address these predictions.

Chapter 2 describes the first version of a cognitive ACT-R model that simulates the acquisition of object pronouns. On the basis of this model I predict that adult-like interpretation, but not adult-like production, of object pronouns requires sufficient processing speed. This chapter also presents a behavioral experiment with 4-6 old Dutch-speaking children testing this prediction. The experiment shows that children who show a Delay of Principle B Effect in acquisition perform better in pronoun comprehension if they are presented with slowed-down speech. This result supports the prediction of our model.

Chapter 3 describes a cognitive model that explains the acquisition of subject pronouns in discourse. The model is an extension of the cognitive model presented in Chapter 2. The assumptions and implications of the model are discussed and the chapter concludes with specific and testable predictions. On the basis of this model, we argue that adult-like interpretation and production of subject pronouns in discourse requires sufficient working memory (WM) capacity and sufficient speed of processing.

Chapter 4 addresses the predictions from the cognitive model presented in Chapter 3. We present a dual-task experiment with adult participants in which we manipulated WM load. The behavioral results indicate that with higher WM load, adults show more child-like performance on their interpretation of subject pronouns in discourse.

Chapter 5 presents a follow-up study of the dual-task experiment in Chapter 4. ERPs were measured during a similar dual-task experiment. The results indicate that WM load and discourse structure influence the cognitive representation of the discourse structure already very early in the discourse, disambiguating potentially ambiguous pronouns before they are encountered.

The cognitive model simulating subject pronoun processing is an extension of the earlier implemented model simulating object pronoun processing. The only addition is a component of discourse processing. This raises the question whether discourse also influences object pronoun processing. Although we did not simulate influences of the linguistic discourse on object pronoun processing, we predict on the basis of our earlier model that, whereas off-line performance of object pronoun interpretation is not expected to be affected in adults, their on-line processing may be affected. Chapter 6 presents a pupil dilation study with Dutch adults, in which we investigated whether discourse affects the on-line processing of object pronouns. Our results show very early effects of linguistic discourse, and also of visual context, suggesting that on-line processing of object pronouns is influenced by the referents' discourse saliency.

Chapter 7 evaluates the cognitive models on the basis of the empirical results. The chapter concludes with the implications of this novel account of the interpretation and production of pronouns in discourse for language acquisition, language processing and language disorders.

Notes

[1] Mart Smeets is a long-time sports presenter and commentator with the Dutch public broadcaster (see http://en.wikipedia.org/wiki/Mart_Smeets). Epke Zonderland is a Dutch gymnast and 2012 Olympic gold medalist in the high bar (see http://en.wikipedia.org/wiki/Epke_Zonderland).

chapter 2

Cognitive architectures and language acquisition:

A case study in pronoun comprehension

This paper was previously published as:
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.

Abstract

In this paper we discuss a computational cognitive model of children's poor performance on pronoun interpretation (the so-called Delay of Principle B Effect, or DPBE). This cognitive model is based on a theoretical account that attributes the DPBE to children's inability as hearers to also take into account the speaker's perspective. The cognitive model predicts that child hearers are unable to do so because their speed of linguistic processing is too limited to perform this second step in interpretation. We tested this hypothesis empirically in a psycholinguistic study, in which we slowed down the speech rate to give children more time for interpretation, and in a computational simulation study. The results of the two studies confirm the predictions of our model. Moreover, these studies show that embedding a theory of linguistic competence in a cognitive architecture allows for the generation of detailed and testable predictions with respect to linguistic performance.

1 Introduction

An influential but also controversial distinction in linguistic research is the distinction between linguistic competence and linguistic performance (Chomsky, 1965). Linguistic competence pertains to the idealized linguistic knowledge a language user has of his or her language, which is often contrasted with linguistic performance, the actual use of this knowledge in concrete situations. This distinction between competence and performance provided a rationale for studying linguistic phenomena separately from cognitive factors. However, this distinction also created the methodological problem that it became impossible to empirically test theories of linguistic competence solely by studying linguistic performance. As a result, linguistic analyses appealing to aspects of linguistic performance such as insufficient working memory capacity, processing limitations or pragmatic skills are difficult to evaluate. Nevertheless, such analyses have been proposed in many areas of language acquisition to explain differences in linguistic performance between children and adults.

The aim of this paper is to show that embedding a theory of linguistic competence in a cognitive architecture may allow for the generation of detailed and testable predictions with respect to linguistic performance. A cognitive architecture is a general framework that incorporates built-in and well-tested parameters and constraints on cognitive processes. Within a cognitive architecture, computational models can be built that simulate the cognitive processes involved in performing a task such as interpreting a sentence. The predictions generated by these computational models can be tested on the basis of empirical data, for example the performance results obtained from a psycholinguistic experiment. As a case study, we present an account of the Delay of Principle B Effect in language acquisition (e.g., Chien & Wexler, 1990; Jakubowicz, 1984; Koster, 1993). The Delay of Principle B Effect (DPBE) concerns the observation that children's comprehension of pronouns is delayed in comparison with their comprehension of reflexives. Initially, children show incorrect performance on pronoun comprehension as well as on reflexive comprehension. However, when they have mastered reflexive comprehension, they still show incorrect performance on pronoun comprehension. This phenomenon in language acquisition is referred to as the DPBE. It can take several years before children show correct performance on both pronoun comprehension and reflexive comprehension.

The DPBE has received a variety of explanations, many of which appeal to performance factors to account for children's errors in comprehending pronouns. One such explanation is formulated within the linguistic framework of Optimality Theory (Hendriks & Spenader, 2006). We show how a cognitive model can be built within the cognitive architecture ACT-R (Anderson et al., 2004) that implements an optimality theoretic explanation of the DPBE. The resulting cognitive model predicts that children will make fewer errors in their interpretation of pronouns but not in their interpretation of reflexives if they are given more time for comprehension, for example by slowing down the speech rate. We tested this prediction empirically in a psycholinguistic study as well as in a computational simulation study.

The organization of this paper is as follows. First, we discuss the DPBE and several of the proposed explanations to account for this delay in language acquisition, including a detailed

account of the optimality theoretic explanation of the DPBE. Then we present a cognitive model that is based on the optimality theoretic explanation of the DPBE. The hypotheses derived from this cognitive model are first tested in a psycholinguistic experiment involving 75 Dutch children between 4;1 and 6;3 years old. Then a simulation study is discussed in which the effects of speech rate on the comprehension of sentences with pronouns and reflexives are modeled. In this second study, the performance of a group of children is simulated and compared to the results of the psycholinguistic experiment. The paper concludes with a discussion of the considerations and limitations in using cognitive models to study theories of language acquisition.

2 Delay of Principle B Effect (DPBE)

A well-established finding in language acquisition research is the observation that, in languages such as English, French and Dutch, children's comprehension of pronouns is delayed in comparison with their comprehension of reflexives (e.g., Chien & Wexler, 1990; Jakubowicz, 1984; Koster, 1993; Philip & Coopmans, 1996; Spenader et al., 2009). This phenomenon is called the Delay of Principle B Effect (DPBE). Principle B is one of the two principles of Binding Theory that relate to the adult use and interpretation of reflexives and pronouns (Chomsky, 1981):

1a. Principle A: a reflexive must be bound in its local domain.

1b. Principle B: a pronoun must be free in its local domain.

The local domain is defined as the minimal clause containing both the lexical anaphor and a subject. An anaphor is bound when it is co-indexed with and c-commanded by an antecedent.[1] Sentences 2a and 2b illustrate the application of Principles A and B:

2a. The penguin$_i$ is hitting himself$_{i/*j}$ with a pan.

2b. The penguin$_i$ is hitting him$_{*i/j}$ with a pan.

The reflexive *himself* in Sentence 2a can only co-refer with the local subject the penguin, in accordance with Principle A, and may not co-refer with another referent. In contrast, Principle B prevents the pronoun *him* in Sentence 2b from co-referring with the penguin. Therefore, him must co-refer with another referent present in the linguistic or extra-linguistic context. From the age of 3;0 on, children are able to interpret sentences with reflexives, like 2a, correctly, thus displaying knowledge of Principle A. However, up to the age of 6;6, children show difficulties in the interpretation of pronouns in sentences like 2b (e.g., Chien & Wexler, 1990). They seem to choose freely between a coreferential interpretation, in which the pronoun co-refers with the local subject, and a disjoint interpretation, in which the pronoun co-refers with an antecedent outside its local domain. Thus, in comprehension children act as if they only have access to Principle A. Their acquisition of Principle B seems to be delayed.

2.1 Explanations of the DPBE

To explain the Delay of Principle B Effect (DPBE), several theories have been proposed within a nativist framework (a notable exception is the usage-based account of Matthews et al., 2009). In this section, we limit ourselves to two well-accepted theories: the pragmatic account of Thornton and Wexler (1999), and the processing account of Reinhart (2006). Both Thornton and Wexler's and Reinhart's account proceed from a nativist view on language. Hence, they assume that children have knowledge of both Principle A and B, and should in principle be able to apply this knowledge. However, the accounts differ in their explanation of why Principle B is delayed.

Thornton and Wexler (1999) propose that the DPBE is caused by a deficiency in pragmatic knowledge. The starting point for their theory is the observation that in certain special contexts a pronoun may receive a coreferential interpretation, for example when the event being described is unexpected or uncharacteristic. To indicate that such an exceptional coreferential interpretation is intended, speakers stress the pronoun ('Mama Bear is washing HER', see Thornton & Wexler, 1999, p. 94), in addition to providing special pragmatic context. Thornton and Wexler argue that children do not yet have sufficient world knowledge and pragmatic knowledge to determine whether the event described by the sentence reflects a typical or atypical situation, that is, to evaluate whether the context licenses a coreferential interpretation. Furthermore, Thornton and Wexler argue that children do not recognize stress on a pronoun as an indication that the speaker intended to express an atypical interpretation. As a result, children accept a coreferential interpretation of a pronoun sentence such as 'Mama Bear is washing her'. For adult language users, only a disjoint interpretation is possible for this sentence, because adults do not allow a coreferential interpretation in the absence of stress. So children over-accept coreferential interpretations of pronouns because they are unable to distinguish the contexts that license coreferential interpretations from the contexts that do not license such interpretations. Children will have to acquire the world knowledge and pragmatic knowledge necessary to disallow a coreferential interpretation of a pronoun in non-exceptional contexts.

Although their account focuses on the comprehension of pronouns, Thornton and Wexler point out that this lack of pragmatic knowledge has ramifications for children's production as well (1999, p. 95). However, under their account it remains a mystery why children who show difficulties on pronoun comprehension at the same time show adult-like performance on pronoun production (see De Villiers et al., 2006; Spenader et al., 2009).

In contrast to Thornton and Wexler, Reinhart (2006) argues that children possess all knowledge required for the interpretation of pronouns. The crucial difference between children and adults is that children fail to complete the operation of *reference-set computation*. Reference-set computation is an operation that is performed by the parser to choose between multiple interpretations generated by the grammar. The operation is required for determining whether a coreferential interpretation is permitted for a pronoun. For a sentence such as 3, for example, the grammar generates two different derivations: one giving rise to a bound variable interpretation 3a, and one giving rise to a coreferential 3b or disjoint 3c interpretation. A coreferential interpretation arises if the two variables *x* and *y* both happen to be

resolved to the same referent, in this case Lili, whereas a disjoint interpretation arises if *x* and *y* are resolved to different referents.

3. Only Lili thinks she's got the flu. (adapted from Reinhart, 2006, p. 167)
    a *Bound variable interpretation:* Only Lili (λx (*x* thinks *x* has got the flu))
    b *Coreferential interpretation:* Only Lili (λx (*x* thinks *y* has got the flu) & *y* = Lili)
    c *Disjoint interpretation:* Only Lili (λx (*x* thinks *y* has got the flu) & $y \neq$ Lili)

The grammar allows the bound variable interpretation 3a for Sentence 3, because the pronoun *she* is not bound within its local domain (cf. Principle B). The grammar also allows the pronoun to be interpreted as a free variable, giving rise to the disjoint interpretation 3c. Whether coreferential interpretation 3b is allowed, however, must be determined through reference-set computation. Reference-set computation involves the comparison of pairs of derivations and their corresponding interpretations. A coreferential interpretation is allowed for 3 only if this interpretation is different from the bound variable interpretation. If these interpretations are indistinguishable, a coreferential interpretation is not allowed because it is inefficient to revert back to an interpretation that is ruled out by the grammar through the discourse option of coreference. With respect to sentence 3, the coreferential interpretation is allowed, because 3a and 3b have slightly different meanings. Interpretation 3a entails that other people do not think that they have got the flu, whereas interpretation 3b entails that other people do not think that Lili has got the flu. The situation is slightly different for the sentence in 4.

4. Mama Bear is washing her.
    a *Bound variable interpretation:* Mama Bear (λx (*x* is washing *x*))
    b *Coreferential interpretation:* Mama Bear (λx (*x* is washing *y*) & *y* = Mama Bear)
    c *Disjoint interpretation:* Mama Bear (λx (*x* is washing *y*) & $y \neq$ Mama Bear)

For this sentence, the grammar (Principle B) disallows the bound variable interpretation 4a, because the pronoun *her* would be bound within its local domain. Although 4a is disallowed by the grammar, reference-set computation nevertheless requires that a bound variable derivation is constructed and its interpretation is compared with the coreferential interpretation 4b. Because the two interpretations are indistinguishable, the coreferential interpretation is not allowed for sentence 4. Consequently, only the disjoint interpretation 4c is possible for this sentence.

Reinhart argues that children may be unable to perform this operation of reference-set computation because of working memory limitations. If children fail to complete the operation of reference-set computation, they resort to a guessing strategy and arbitrarily choose between a coreferential and a disjoint interpretation. Other strategies are conceivable as well and are used with other marked forms requiring reference-set computation, such as contrastive stress. Only when children have developed sufficient working memory capacity will they be able to complete the operation of reference-set computation and disallow the coreferential interpretation for pronouns. Because the grammar generates two derivations for pronoun sentences but not for reflexive sentences, reference-set computation is not involved in

the interpretation of reflexives. With respect to the production of pronouns, as speakers know which meaning they intend for the utterance, reference-set computation is not involved in production either. This would explain why children are able to produce pronouns correctly from a young age on while still having difficulties with the comprehension of pronouns (De Villiers et al., 2006; Spenader et al., 2009).

In the next section, we contrast these theories with an alternative theory: the optimality theoretic account of Hendriks and Spenader (2006), which assumes that only Principle A is part of grammar and Principle B is a derived effect.

3 Optimality Theory explanation of the DPBE

A third type of explanation of the DPBE is provided by Hendriks and Spenader (2006). They argue that the DPBE is the result of a direction-sensitive grammar, that is, a grammar that has different effects in production and comprehension. Their account is formulated within the framework of Optimality Theory (OT), a linguistic framework that models the relationship between a surface form and its underlying structure by means of optimization from a particular input to the optimal output for that input (Prince & Smolensky, 2004). In the domain of semantics, OT describes the relation between an input form and the optimal meaning for that form (e.g., Hendriks & de Hoop, 2001). Applied to syntax, OT describes the relation between an input meaning and the optimal form for expressing that meaning. OT thus provides an account of linguistic competence with respect to language production (i.e., OT syntax) as well as language comprehension (i.e., OT semantics). In OT, the grammar consists of a set of violable constraints, rather than inviolable rules. For every input, which can be either a form or a meaning, a set of potential outputs, or candidates, is generated. These candidates are evaluated on the basis of the constraints of the grammar. In OT, constraints are as general as possible and hence may conflict. OT resolves conflicts among constraints by ranking the constraints in a language specific hierarchy on the basis of their strength. One violation of a stronger (i.e., higher ranked) constraint is more important than many violations of a weaker (i.e., lower ranked) constraint. The optimal candidate is the candidate that commits the least severe constraint violations. Only the optimal candidate is realized.

3.1 Direction-sensitive grammar

For their explanation of the DPBE, Hendriks and Spenader (2006) exploit the fact that an OT grammar is inherently direction-sensitive: The form-meaning relations defined by the OT grammar are not necessarily the same from the speaker's perspective (involving optimization from meaning to form) as from the hearer's perspective (involving optimization from form to meaning) (Smolensky, 1996). This property of OT is a result of the output orientation of the markedness constraints in OT. OT assumes two kinds of constraints. *Faithfulness* constraints evaluate the similarity between input and output. Because faithfulness constraints pertain to the mapping between input and output, these constraints are direction-insensitive and also apply in the reverse direction of optimization. An example is the constraint PRINCIPLE A (5),

which prohibits reflexives from being locally free. This constraint induces hearers to assign a locally bound interpretation to reflexives and at the same time prohibits speakers to express a disjoint interpretation by using a reflexive.

5. PRINCIPLE A: A reflexive must be bound in its local domain.

Markedness constraints on forms, on the other hand, reflect a preference for unmarked forms, irrespective of their meaning. Because they pertain to the output only, markedness constraints on form only have an effect when a form must be selected from a set of candidate forms. That is, they only have an effect from the speaker's perspective. An example is the constraint AVOID PRONOUNS. For hearers, this constraint does not have any effect, because for hearers the form is already given as the input. The hearer's task is to select the optimal meaning for this form. Since the constraint AVOID PRONOUNS does not distinguish between potential meanings, it does not have any effect from the hearer's perspective. The constraint AVOID PRONOUNS is part of the constraint hierarchy REFERENTIAL ECONOMY (6). This constraint hierarchy consists of several markedness constraints, of which AVOID REFLEXIVES is the lowest ranked. The hierarchy reflects a preference for less referential content: Reflexives are preferred over pronouns, and pronouns over full NPs.

6. REFERENTIAL ECONOMY: Avoid full NPs $\gg$ Avoid pronouns $\gg$ Avoid reflexives[2].

In this discussion we limit ourselves to the choice between pronouns and reflexives and hence only consider the constraint AVOID PRONOUNS. This constraint is violated by any pronoun in the output, and is satisfied by any reflexive in the output. The presence of markedness constraints such as AVOID PRONOUNS can lead to an asymmetry between production and comprehension, as is shown below.

The evaluation of candidates on the basis of the constraints of the grammar can be illustrated with an OT tableau. Figure 1 displays the two comprehension tableaux representing the comprehension of a reflexive and the comprehension of a pronoun, respectively. The input to a comprehension tableau is a form and the output is the optimal meaning for this form. The constraints are presented in columns in order of descending strength, from left to right. PRINCIPLE A must be ranked higher than AVOID PRONOUNS because otherwise pronouns would never be selected. The relevant candidate outputs (in this case, potential meanings for the input form) are listed in the first column. A violation of a constraint is marked with a '*', and a fatal violation with a '!'. The optimal output is marked by '☞'.

When a hearer encounters a pronoun or a reflexive, he has to choose between a coreferential interpretation (first row) and a disjoint interpretation (second row). The coreferential interpretation is the optimal interpretation for a reflexive (Figure 1a), because the disjoint interpretation violates the strongest constraint PRINCIPLE A, whereas the coreferential interpretation satisfies this constraint. When comprehending a pronoun (Figure 1b), PRINCIPLE A is not relevant because it does not define the antecedent possibilities of pronouns. Because AVOID PRONOUNS does not apply in comprehension, both the coreferential interpretation and the disjoint interpretation are optimal candidates according to the grammar. As a result, pronouns are ambiguous. Hence, children might randomly choose one of the two candidate meanings when no contextual clues are available.

	Input: reflexive	Principle A	Avoid Pronouns
☞	**coreferential**		
	disjoint	*!	

Figure 1: a. OT tableau of the comprehension of a reflexive

	Input: pronoun	Principle A	Avoid Pronouns
☞	**coreferential**		
☞	**disjoint**		

Figure 1: b. OT tableau of the comprehension of a pronoun.

Figure 2 shows the tableaux for the production of a coreferential interpretation and a disjoint interpretation, respectively. A violation of a constraint is marked with a '*', and a fatal violation with a '!'. The optimal output is marked by '☞'.

	Input: coreferential	Principle A	Avoid Pronouns
☞	**reflexive**		
	pronoun		*!

Figure 2: a. OT tableau of the production of a coreferential meaning

	Input: disjoint	Principle A	Avoid Pronouns
	reflexive	*!	
☞	**pronoun**		*

Figure 2: b. OT tableau of the production of a disjoint meaning.

When a speaker wishes to express a coreferential meaning (Figure 2a), the relevant competing candidate forms are a pronoun and a reflexive. Principle A does not distinguish between these two candidates, because this constraint allows a coreferential interpretation to be expressed by a reflexive as well as a pronoun. However, Avoid Pronouns prefers reflexives over pronouns. Therefore, a reflexive is a better form for expressing a coreferential meaning than a pronoun. On the other hand, a pronoun is the optimal form for expressing a disjoint meaning (Figure 2b), because it satisfies Principle A, whereas a reflexive does not.

In summary, an optimality theoretic grammar is direction-sensitive, because the optimal form-meaning pairs in production are not necessarily the same as the optimal form-meaning pairs in comprehension. Specifically, a pronoun can have a coreferential and a disjoint interpretation according to the grammar, whereas the best form for expressing a coreferential interpretation is a reflexive and the best form for expressing a disjoint interpretation is a pronoun according to the same grammar. This fits the pattern typically displayed by 4- to 7-year

old English and Dutch children, leading to an asymmetry between their production and their comprehension (De Villiers et al., 2006; Spenader et al., 2009).

3.2 Explanation for adults' comprehension of pronouns

In contrast to children, adult language users always interpret a pronoun as having a disjoint meaning. According to Hendriks and Spenader (2006), the difference between children and adults is that adult hearers also take into account the perspective of the speaker, whereas children only consider their own perspective. The adult way of interpretation can be modeled as bidirectional optimization (Blutner, 2000). Figure 3 illustrates the serial implementation of bidirectional optimization proposed by Hendriks, van Rijn, and Valkenier (2007).

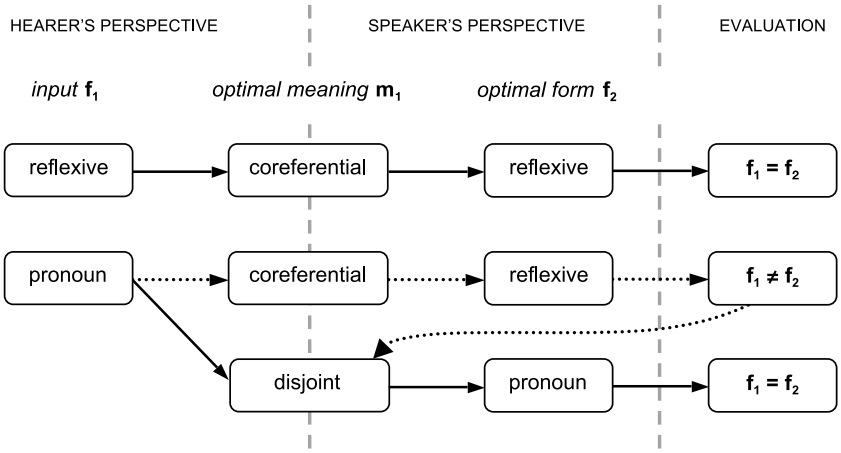


Figure 3: Taking into account the speaker's perspective in comprehension. The coreferential interpretation for pronouns (represented by the dotted line) is blocked because a coreferential interpretation is best expressed by a reflexive.

When an adult hearer encounters a pronoun or a reflexive, he has to determine the optimal meaning for this form. This requires the hearer to optimize in the hearer's direction of optimization: from form to meaning. In addition, however, the hearer must also check whether the selected meaning is indeed expressed by the encountered form. This requires that the hearer also optimizes from meaning back to form, that is, that the hearer adopts the speaker's perspective. When comprehending reflexives, this process of bidirectional optimization leads to the same result as unidirectional optimization. In both cases, the optimal meaning for a reflexive is a coreferential interpretation. However, when comprehending pronouns, bidirectional optimization leads to a different result. Recall that, from a hearer's perspective, pronouns are ambiguous and can also receive a coreferential interpretation. From the speaker's perspective, however, a coreferential meaning is best expressed using a reflexive. If a hearer were to select the coreferential meaning for the ambiguous pronoun in the first

step of optimization, he would find out in the second step of optimization that a coreferential meaning is best expressed with a reflexive. So the resulting form (a reflexive) is different from the encountered form (a pronoun). As a consequence, the coreferential interpretation is blocked for pronouns, and pronouns are only assigned a disjoint interpretation.

In summary, children's pattern of comprehension and production can be explained by unidirectional optimization, which is a formalization of the idea that children only consider their own perspective. The adult pattern can be explained by bidirectional optimization, which is a formalization of the idea that adults take into account the opposite perspective in addition to their own perspective. This OT explanation provides an adequate description of the Delay of Principle B Effect. It can account for the observation that the interpretation of pronouns is acquired later than the interpretation of reflexives. It also explains why children's production of pronouns may already be adult-like, while their comprehension of pronouns is still poor. Furthermore, the analysis of the DPBE can be generalized to other acquisition delays, either in comprehension or in production. This contrasts with the processing account of Reinhart (2006), which only predicts delays in comprehension.

4 Testing linguistic theories

In the previous sections, we discussed three different explanations for the DPBE: Thornton and Wexler's (1999) pragmatic account, Reinhart's (2006) processing account and Hendriks and Spenader's (2006) OT account. These explanations illustrate the lack of consensus with respect to the cause of the DPBE. An important reason for this lack of consensus is that it is difficult to contrast the theories on the basis of linguistic data alone. The theories mentioned above attribute the DPBE to non-linguistic factors such as a lack of pragmatic knowledge, limited working memory capacity or the inability to take into account another person's perspective. However, without further specification of these non-linguistic factors and how they influence linguistic performance, it is difficult to evaluate these theories. To arrive at a full understanding of linguistic competence, it is therefore essential that theories of linguistic competence are tested in combination with viable theories of pragmatic reasoning, memory, parsing and other cognitive processes. Only then will it be possible to generate precise predictions for linguistic performance that can be empirically evaluated on the basis of experimental data.

A possible way to combine a theory of linguistic competence with theories of cognition and cognitive processes is by embedding the linguistic theory in a cognitive architecture. Cognitive architectures combine several theories of different cognitive subsystems into a single theory of the human cognitive system. A number of architectures have been proposed e.g., EPIC: Meyer & Kieras, 1997; Soar: Newell, 1990; ACT-R: Anderson et al., 2004 that offer a computational environment in which models can be constructed of the phenomena under study. By constructing a model in the context of an architecture, the model automatically respects the assumptions of the architecture.Computational simulations are a powerful tool for testing theories since they allow for assessing the completeness of the theoretical account. Also, they make explicit which cognitive processes are required for explaining the phenomenon

that is studied. The output of a simulation typically consists of the observed behavior and of estimates of the time it takes to perform the task. Therefore, precise predictions can be generated of human behavior (for a review of language-related computational models, see Dijkstra & De Smedt, 1996).

As we saw in the previous section, OT provides a way to account for children's and adults' linguistic competence with respect to pronouns. However, since OT is a theory of linguistic competence, it does not provide an explanation for the change in optimization mechanism between children and adults. Also, OT does not make any predictions about the time it takes to develop the ability to apply bidirectional optimization, or about the factors that are relevant in developing this ability.

The following section presents a computational cognitive model of the acquisition of pronoun comprehension that is based on the theoretical OT model of Hendriks and Spenader (2006) and is implemented in the general cognitive architecture ACT-R (Anderson et al., 2004). In the model, ACT-R interacts directly with the OT grammar to produce linguistic performance (cf. Hendriks et al., 2007; see Misker & Anderson, 2003, for an alternative approach to integrating OT and ACT-R). This is possible because of two important properties of OT: its robustness and its cross-modularity. Because OT is robust and does not pose any restrictions on the input, OT is able to assign an optimal output even to incomplete, dispreferred or ill-formed inputs. Hence, it is able to explain incremental parsing and certain parsing preferences without having to assume a separate parser (e.g., S. Stevenson & Smolensky, 2006). Furthermore, because OT can be applied to any linguistic domain, OT constraints can be ordered in one large constraint hierarchy. As a consequence, an OT grammar is inherently cross-modular and does not require any interfaces to mediate between different linguistic modules. These two properties allow us to implement the OT grammar (i.e., the constraints and their ranking) directly into the cognitive architecture ACT-R. As we will show, the resulting computational cognitive model generates testable predictions with respect to children's and adults' performance on pronoun comprehension.

5 Cognitive models of language acquisition

The computational cognitive model we constructed is built in the cognitive architecture ACT-R (Adaptive Control of Thought-Rational; Anderson et al., 2004). ACT-R is both a theory of cognition and a modeling environment. As a theory of cognition, its aim is to explain human cognition and to account for a broad range of data from psychological and neurocognitive experiments.[3] It has a modular structure: Each of ACT-R's modules is based on smaller theories on cognition. For example, ACT-R contains a theory about retrieving declarative knowledge that is based on Anderson and Schooler's rational analysis of memory (Anderson & Schooler, 1991) and a theory about the processing of auditory stimuli that is loosely based on EPIC (Meyer & Kieras, 1997). ACT-R also is a modeling environment that can be used to implement a computational simulation of a specific task. The architecture constrains these simulation models to ensure psychological plausibility. The constraints imposed on the models are based on experimental data and define how information is processed, stored and retrieved

within modules, and how information is communicated between modules (Anderson, 2007). Although many decisions have to be made when a linguistic analysis is translated into a computational simulation, mainly related to the non-linguistic aspects of the task, the constraints of the cognitive architecture guide these decisions.

5.1 General structure of ACT-R

ACT-R distinguishes several modules that are involved in different aspects of human cognitive functioning. The two main modules of ACT-R are declarative memory and the central production system. Declarative memory contains chunks of factual information. The central production system contains IF-THEN rules. The IF-clause of each production rule specifies a number of conditions that must be met for that production rule to be considered for execution. For example, a production rule that initiates a search in memory for alternative interpretations of a linguistic input is subject to the condition that a linguistic input is available in memory that has not already been fully processed, and that the memory system is currently not in use by another operation. The THEN-clause specifies which actions need to be performed if that production rule is selected for execution (for example, the instruction to initiate the retrieval of a memory element, or to initiate a key press). At each time step, the central production system matches the production rules to the current state of the system, and the most active matching rule is selected for execution. The activation value of production rules reflects the utility of that rule and is an expression of the expected benefits of executing that production rule discounted for the costs associated with that production rule. Elements in declarative memory are ranked on the basis of their activation value. The activation value of declarative memory elements (often called chunks) reflects the usefulness of that chunk in the current context. This activation value is based on a weighted average of the number of prior occurrences of that chunk in general, and the number of prior occurrences of that chunk in the current context.

An assumption of ACT-R that is important for the present study is the assumption that every operation, for example the retrieval of a fact from declarative memory or the execution of a production rule, takes a certain amount of time. The total execution time of a cognitive process is not simply the sum of the durations of all constituting operations, as the different modules can operate in parallel. However, each module in itself can only perform a single action at a time. Thus, the duration of a process critically depends both on the timing of the serial processes within a module, and on how the different modules interact. To provide specific time estimations for a cognitive process, a computational simulation model can be constructed within the ACT-R system that provides precise predictions when it is run (Anderson et al., 2004).

5.2 Modeling unidirectional optimization

In this section, we present an ACT-R model that implements Hendriks and Spenader's (2006) theoretical account of the DPBE. Our computational DPBE/ACT-R model is a refined implementation of Hendriks et al. (2007), that enables us to derive more precise predictions re-

lated to the DPBE[4]. The main difference between our implementation and the original model (Hendriks et al., 2007) is that our DPBE/ACT-R model is more generic than the original model. Our DPBE/ACT-R model can simulate not only the acquisition of pronoun comprehension, but the acquisition of pronoun production as well. Moreover, the current implementation allows for more principled timing of the processes involved in linguistic performance (see R. L. Lewis & Vasishth, 2005; van Maanen, van Rijn, & Borst, 2009; van Rijn & Anderson, 2003, for other approaches modeling temporal aspects of linguistic processing in ACT-R).

In our model, different candidate forms and candidate meanings are implemented as chunks in declarative memory. From a hearer's perspective, there are two possible candidate interpretations for a pronoun or a reflexive: a disjoint meaning and a coreferential meaning. From a speaker's perspective there are two possible candidate forms to express a disjoint meaning or a coreferential meaning: a pronoun or a reflexive. These four different candidates are represented as separate chunks. The optimality theoretic constraints are implemented in terms of the violations they incur. Each constraint is represented as a collection of chunks. The chunks specify for each possible input which candidate outputs violate this constraint. As there are four possible inputs in the current domain, each constraint is represented as four chunks. Production rules define strategies to retrieve forms, meanings, and constraints from memory. Figure 4 illustrates the process of finding the optimal meaning. Although the process can be applied to comprehension as well as production, in this study we focus on comprehension.

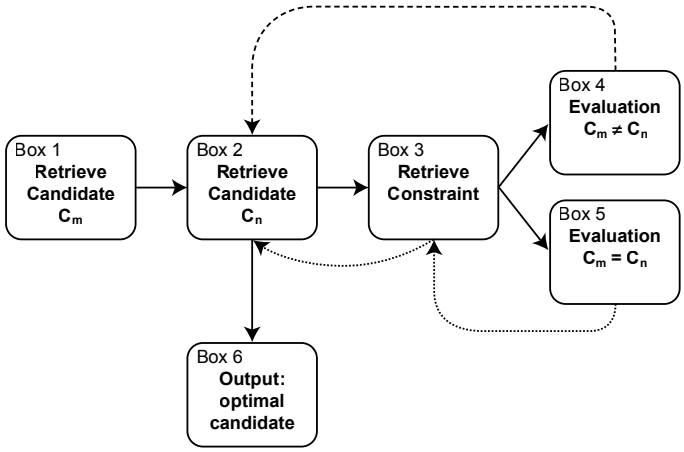


Figure 4: Structure of the ACT-R model of learning to optimize bidirectionally (adapted from Hendriks et al., 2007).

In comprehension, the input for the DPBE/ACT-R model consists of a pronoun or a reflexive for which the optimal interpretation has to be determined. The first step is to retrieve two of the possible candidates from declarative memory. After the retrieval of the two can-

didates, a chunk is retrieved representing a constraint that can be used to evaluate the two candidates. Because chunks are ordered based on their activation value, the system will retrieve the highest ranked constraint first. If one of the candidates violates the constraint, that candidate is replaced by another candidate from declarative memory, and another process of comparison takes place. If there is no other candidate, the remaining candidate will be selected as the optimal meaning. If the two candidates show the same pattern of constraint violations (both violating or satisfying this constraint), the next constraint will be retrieved. If none of the constraints distinguishes between the two candidate meanings, then one of the meanings is randomly selected as the optimal meaning. This process is similar to a recursive optimization process that finds the optimal candidate by evaluating the candidates against the highest ranked constraint and evaluating the candidates against lower ranked constraints only if necessary. A complete recursive optimization process would continue until all potential candidates are evaluated. However, the optimization process can be interrupted in the simulations because of cognitive constraints.

In the optimality theoretic analysis of Hendriks and Spenader (2006), PRINCIPLE A is the strongest of the two constraints. When the input is a reflexive, the application of PRINCIPLE A is already sufficient to select the coreferential meaning as the optimal meaning. However, when the input is a pronoun, the application of both PRINCIPLE A and AVOID PRONOUNS is insufficient to distinguish between the two candidate meanings. Therefore, one of these candidates is randomly selected as the optimal meaning. At this stage, the model performs at chance on pronoun comprehension, whereas it shows almost correct performance on reflexive comprehension. Because the model requires more steps to arrive at an interpretation for the pronoun (first applying two constraints and then selecting a candidate at random) than for the reflexive (just applying one constraint), it is predicted that it takes the model more time to process a pronoun than a reflexive.

5.3 Modeling bidirectional optimization

Bidirectional optimization can be thought of as two processes of unidirectional optimization to be performed during on-line sentence processing (Hendriks et al., 2007). In a processing account of bidirectional OT, a straightforward implementation of bidirectional optimization is to have the second step of optimization follow the first step of optimization, as the second step requires the output of the first step. Therefore, we have implemented bidirectional optimization as two unidirectional optimization processes that are performed in sequence. This was already schematically described in Figure 3.

In the computational simulation new inputs arrive at a fixed rate. As in the situation in which an external speaker determines the speaking rate, the model as a hearer cannot influence this rate. Therefore, the amount of time available for selecting an optimal meaning is limited. Because bidirectional optimization consists of two sequential processes of unidirectional optimization, bidirectional optimization takes more time than unidirectional optimization. In the DPBE/ACT-R model, initially the model can only perform a single process of unidirectional optimization within the limited time and is therefore unable to perform bidi-

rectional optimization during on-line sentence processing. This results in performance that is similar to the performance of young children. As discussed earlier, a unidirectional process results in the correct interpretation of reflexives. However, because two steps of unidirectional optimization are needed for correct pronoun interpretation, performance on pronouns remains at chance level. Only if more time is provided, or when processing efficiency improves, bidirectional optimization becomes possible. This account of the DPBE implies that children are in principle capable of applying bidirectional optimization but do not succeed because of limited resources.

In the ACT-R architecture, higher processing efficiency is obtained through the mechanism of production compilation (Taatgen & Anderson, 2002). Production compilation is a learning mechanism that combines two production rules that are repeatedly executed in sequence into one new production rule. By means of this learning mechanism cognitive processing becomes much faster, since the new rule has the same functionality as the two production rules before compilation. For example, the model contains two production rules that retrieve candidates from declarative memory. The first production rule requests the retrieval of a candidate on the basis of the received input. The next production rule processes that retrieval and requests another candidate that is not the same as the first retrieved candidate. After repeatedly using these two rules in sequence, the production compilation mechanism creates a new production rule that stores the information of the two candidates at the same time. As this production rule is much more efficient than the two original production rules, this new rule will be preferred by the model. This new rule can again be combined with other production rules in exactly the same way. For example, a production rule may be created that not only retrieves two candidates, but also evaluates these two candidates on the basis of the highest ranked constraint. Eventually, sentence processing is performed fast enough for bidirectional optimization to succeed within the available amount of time. Note that the time course of learning depends on the frequencies of the input forms, as the compilation of production rules is a function of the number of times a set of rules has been executed in sequence.

In summary, we modeled bidirectional optimization as two sequential processes of unidirectional optimization. If the model cannot perform both steps within the allotted time, pronouns remain ambiguous and a guessing pattern emerges. However, when the model is given more time for interpretation, it will show increased performance on pronoun comprehension. To test this prediction, we performed a psycholinguistic study described in the next section.

6 Study 1: Experimental study

In this section, we present the results of a psycholinguistic study that we carried out to test the predictions of the cognitive model discussed in the previous section. Based on the properties of the DPBE/ACT-R model, we predict that performance of children displaying the DPBE increases when they are given more time for interpretation. We allowed children more time for interpretation by slowing down the speech rate. In contrast to the predicted increase

in performance on pronoun interpretation, we predict children's performance on reflexive interpretation to remain level.

6.1 Method and materials

A Truth Value Judgment Task was carried out to test children's comprehension of pronouns and reflexives in Dutch. Participants were shown a picture on a computer screen (see Figure 5), and had to judge whether a prerecorded sentence presented to them was a correct description of the picture.

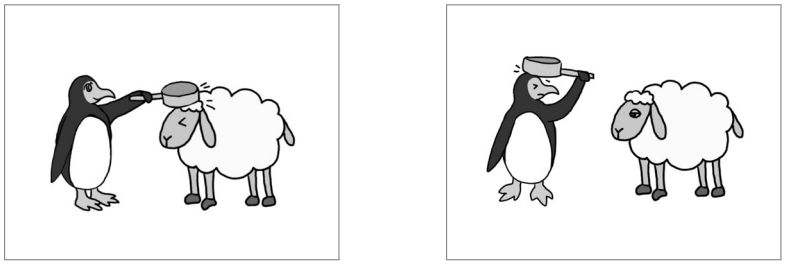


Figure 5: An example of a picture showing a non-reflexive action (left) and a reflexive action (right).

All pictures contained two animals, one of which was depicted as the actor. Both animals were drawn in approximately the same size, to avoid a difference in saliency that may have influenced earlier experiments (for a discussion, see Elbourne, 2005). Each test sentence contained either the reflexive *zichzelf* 'himself' or the pronoun *hem* 'him':

7. Kijk, een pinguïn en een schaap zijn op de stoep.
   De pinguïn slaat hem / zichzelf met een pan.
   'Look, a penguin and a sheep are on the sidewalk.
   The penguin is hitting him / himself with a pan.'

To allow for the experimental manipulation of processing time, the pronouns and reflexives were always followed by a prepositional phrase. The verbs that were used are *bijten* 'to bite', *kietelen* 'to tickle', *schminken* 'to make up', *wijzen naar* 'to point at', *slaan* 'to hit', *vastbinden* 'to tie up', *zien* 'to see', *schilderen* 'to paint', and *tekenen* 'to draw'. All verbs are typically used for describing an other-directed action, thus avoiding a bias towards a coreferential interpretation (Spenader et al., 2009). The same verbs were used in both speech rate conditions, but the sentences differed in the choice of actors and prepositional objects. Half of the sentences were combined with a matching picture and the other half were presented with a non-matching picture. In addition to the test sentences, four control sentences per condition were included to measure the participants' general performance on the task.

All sentences were prerecorded at normal speech rate (mean speech rate 4.0 syllables per second). Sentences for the Slow Speech Rate condition were then digitally slowed-down,

while keeping the pitch constant. Using the software Adobe Audition 1.5, the audio files were stretched 1.5 times, resulting in a reduction of the speech rate with a factor 2/3 (mean speech rate 2.7 syllables per second) (cf. Love, Walenski, & Swinney, 2009; Montgomery, 2004; Weismer & Hesketh, 1996). Native Dutch speaking adults did not report perceiving the slowed-down sentences as disfluent or unnatural. They described the slowed-down sentences as utterances from a slow speaker. During the experiment, the child participants never commented on the speech rate of the test sentences. So there is no indication that slowing down the sentences resulted in an artificial test situation.

6.2 Procedure

Every participant was tested in normal and slow speech rate condition. The order of conditions was counterbalanced over participants. Participants were tested individually in a room by two experimenters. A laptop was used to present the pictures and the prerecorded sentences. The sentences started half a second after the picture appeared on the screen. The participants were instructed to press a button with a green smiley face when they considered the sentence a correct description of the picture, and a button with an orange frowning face when they thought the sentence was not a correct description of the picture. Before the test phase, participants practiced the task with two trial items that were presented in the same speech rate as the following condition. They could take as much time as needed to give a response and they were allowed to hear the prerecorded sentence once more when they asked for it. The conditions were presented as blocks of 20 sentences, i.e., 8 pronoun sentences, 8 reflexive sentences, and 4 control sentences, with a short break in between the two blocks.

6.3 Participants

Seventy-five children between 4;1 and 6;3 years old were tested. They were all recruited from a Dutch local elementary school. From these 75 children, 13 were excluded from further analysis (4 children were bilingual or nonnative Dutch speakers, 5 did not finish the task, and 4 responded incorrectly to more than two out of eight control items). The data of the remaining 62 children (35 boys and 27 girls), ranging in age from 4;1 to 6;2, were used for statistical analysis.

6.4 Results

Looking at the data of all participants, the percentage of correct interpretations was found to be higher for reflexive sentences than for pronoun sentences (90% for reflexives, 60% for pronouns; repeated-measures ANOVA: $F(1,61)=125.968$, $p<0.001$), replicating the results of earlier studies (a.o., Chien & Wexler, 1990; Spenader et al., 2009). Our main question was whether there is a difference in performance between the two speech rate conditions. Statistical analysis of all data showed no significant effect of speech rate on either pronoun comprehension or reflexive comprehension (repeated measures ANOVA: main effect of Speech Rate $F(1,61)<1$; interaction effect of Speech Rate and Expression $F(1,61)<1$).

However, a closer look at the individual data reveals that a possible effect of speech rate may have been masked, because the participants showed different, and sometimes even opposite, behavior on the task as a function of speech rate. In order to investigate the effect of speech rate on pronoun comprehension in more detail, the participants were classified into different developmental stages on the basis of their performance. This division in groups is crucial for the purposes of our study, as only those children who display the DPBE are predicted to show increased performance with slowed-down speech. First, the criteria used for classifying the participants are described. This is followed by more detailed analyses of the effect of speech rate on the different developmental groups.

6.4.1 Classification of different developmental stages

We divided the 62 participants in our study into three different groups, based on their task behavior.[5] For our classification, we defined (almost) correct performance as more than or equal to 80% correct. Incorrect performance was defined as less than 80% correct.
i) Children who showed incorrect performance on pronouns as well as reflexives at Normal Speech Rate were categorized as belonging to the *Incorrect Performance group* (n=5: 3 boys, 2 girls; age 4;3-4;7; mean 4;5),
ii) Children who showed incorrect performance on pronouns but (almost) correct performance on reflexives at Normal Speech Rate were categorized as the *DPBE group* (n=43: 23 boys, 20 girls; age 4;1-6;2; mean 5;1), and
iii) Children who showed (almost) correct performance on both reflexives and pronouns at Normal Speech Rate were categorized as the *Correct Performance group* (n=14: 9 boys, 5 girls; age 4;2-6;0, mean 5;5).
On the basis of the criteria mentioned above, none of the children showed the fourth conceivable pattern of (almost) correct performance on pronouns but incorrect performance on reflexives. Figure 6 shows the distribution of the ages for the three different groups.

Children's scores were analyzed using (logistic) linear mixed-effect models (Baayen, Davidson, & Bates, 2008; Bates, 2005). This type of analysis is more suited for the data than repeated-measures ANOVAs, because several assumptions for using ANOVAs are not met (see Baayen, 2008, for a discussion on this topic). As our DPBE/ACT-R model starts out from the situation in which knowledge of the linguistic constraints and their ranking is already in place, we will only discuss the results of the DPBE group and the Correct Performance group.

6.4.2 Results of the DPBE group

Figure 7 shows the mean percentage of correct interpretations of the 43 children displaying the DPBE. The left plot presents the mean performance on sentences with pronouns and reflexives. The right plot distinguishes between performance on sentences matching the picture and sentences not matching the picture.

Figure 7 shows a clear difference in performance on match items (Normal Speech Rate: 77% , Slow Speech Rate: 74%) versus mismatch items (Normal Speech Rate: 23%, Slow Speech Rate: 34%), probably caused by a yes-bias (see also Chien & Wexler, 1990; Grimshaw & Rosen,

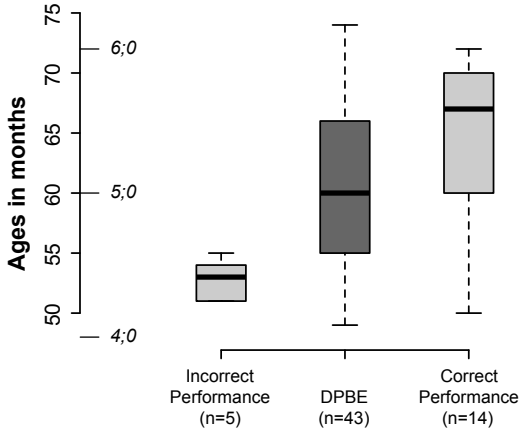


Figure 6: Mean age in months for the different groups in the psycholinguistic study.

1990). To determine the relative contribution of a number of factors on performance, logistic linear mixed-effects models (Bates, 2005) were fit to the data by Laplace approximation. The factors included as fixed effects were: *Block*, a between-subjects factor defining the order of presentation of the two conditions; *Expression*, a within-subjects factor specifying type of anaphor (pronoun or reflexive); *ExpectedAnswer*, a binary within-subjects factor specifying whether the sentence matched the picture or not (yes or no), and a within-subject binary factor *SpeechRate* specifying speech rate (normal or slow). The interactions between Expression, ExpectedAnswer and SpeechRate were included as well. *Subject* and a by-subject effect for ExpectedAnswer were included as random effects, to account for individual differences of the participants and for individual answer biases of participants. Separate sets of models were

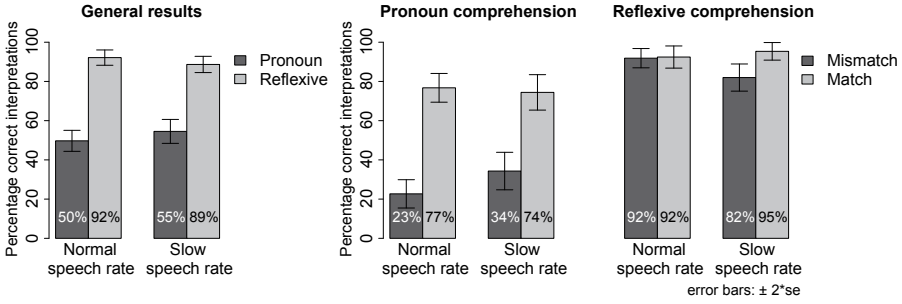


Figure 7: Mean percentage of correct interpretations of sentences with a pronoun or a reflexive in the two speech rate conditions, for the children showing the DPBE (n=43).

constructed with pronouns and reflexives as dependent variables.

For the sentences with *pronouns*, we compared the mixed-effects model that included SpeechRate as a factor with the model that did not, to measure whether manipulation of SpeechRate significantly affected the participants' performance. A comparison was conducted on the basis of the models' log-likelihoods (Baayen, 2008). The comparison showed that the model including SpeechRate explains significantly more variance ($\chi^2(2)= 7.1796$, p=0.028) than the model without SpeechRate. Thus, slowed-down speech has a significant effect on pronoun comprehension. The following factors contributed to the participants' score on the pronoun items: ExpectedAnswer (yes) β =2.964; z = 8.24; p < 0.001, SpeechRate (Slow) β =0.689; z = 2.67; p = 0.008, Block (Slow Speech Rate condition first) β =0.242; z = 0.82; p = 0.412, and the interaction between ExpectedAnswer and SpeechRate β =-0.841; z = -2.25; p = 0.024. The yes-bias, as illustrated in Figure 7 (right plot), is reflected in the significant effect of ExpectedAnswer. The positive β-value of Speech Rate (0.689) indicates that slowed-down speech has a positive effect on pronoun comprehension, although this effect is reduced in the match items, as suggested by the negative coefficient of the interaction effect between Expected Answer and Speech Rate (-0.841). Further analysis of the interaction between ExpectedAnswer and SpeechRate confirmed that there is a significant positive effect of slowed-down speech on the mismatch items (23% correct interpretations in the Normal Speech Rate condition versus 34% correct interpretations in the Slow Speech Rate condition; paired t(42)=2.457, p=0.018). However, no significant difference was found for the match items (paired t(42)<1). The main conclusion from these analyses is that slowed-down speech has a positive effect on pronoun comprehension for children that show the DPBE, as predicted by the DPBE/ACT-R model.

Similar linear mixed-effects models were used to analyze the performance on sentences with *reflexives*. Figure 7 shows almost correct performance on reflexive comprehension with match items as well as mismatch items in the Normal Speech Rate condition. However, in the Slow Speech Rate condition, the percentage of correct responses decreases on the mismatch items, but not on the match items, suggesting a small yes-bias. This decrease in performance on mismatch items also suggests a detrimental effect of slowed-down speech.

Again, we compared a model including the factor SpeechRate with a model without SpeechRate. The model including the factor SpeechRate explains significantly more variance ($\chi^2(2)$=9.757, p=0.008) than the simpler model. Although this shows that slowed-down speech has a significant effect on reflexive comprehension, the effects are not as straightforward as with pronouns. The effects of the included factors on the reflexive items are: Block (Slow Speech Rate condition first) β = -1.696; z = -4.29; p = 0.000, ExpectedAnswer (yes) β = 1.827; z = 2.61; p = 0.009, SpeechRate (Slow) β = -0.967; z = -2.79; p = 0.005, and the interaction between ExpectedAnswer and SpeechRate β = 1.666; z = 2.47; p = 0.013. The negative estimated effect of SpeechRate (-0.967) might be due to interaction between ExpectedAnswer and SpeechRate (1.666). Further analysis revealed that slowed-down speech indeed has a significant effect only in the mismatch (no) items (paired t(42)=-2.418, p=0.020), and not in the match (yes) items (paired t(42) < 1).

In the pronoun analyses, the estimate of Block was not significantly different from zero, but in these reflexive analyses, Block has a negative effect on the percentage of correct inter-

pretations. Children who started the experiment with the Slow Speech Rate condition performed worse on reflexive comprehension in slowed-down speech than children who first participated in the Normal Speech Rate condition. It might be that starting the experiment in the Slow Speech Rate condition triggers other processing strategies, causing additional effects in comprehension. As the effects are more pronounced in the pronoun sentences, a similar effect could be hidden in the variance of that dataset. However, the current data does not allow for testing this.

To summarize, if the child displays the DPBE, slowed-down speech has a positive effect on children's comprehension of pronouns. In contrast, slowed-down speech has a negative effect on children's comprehension of reflexives.

6.4.3 Results of the Correct Performance group

The computational model discussed above predicts that if children are able to take into account both their own perspective and the speaker's perspective under normal conditions, they are also able to do so when they have more time for interpretation. Therefore, the model predicts no effects of speech rate on pronoun or reflexive comprehension in the Correct Performance group. To test this prediction, performance on pronoun comprehension was analyzed, again using linear mixed-effect model comparisons. The factors Block, ExpectedAnswer and SpeechRate were included as fixed effects, as well as the interaction effects of ExpectedAnswer and SpeechRate. In addition, Subject and a by-subject effect for ExpectedAnswer were included as random effects. The factor SpeechRate was found to have a significant effect on pronoun comprehension ($\chi^2(2)= 17.450$, $p<0.001$, with as estimated effect of SpeechRate: β =-1.618; z = -2.99; p = 0.003). In particular, slowed-down speech has a negative effect on pronoun comprehension. Because the effect of SpeechRate is significant both for mismatch items (paired t(13)=-3.647, p=0.003) and match items (paired t(13)=-2.687, p=0.019), slow speech may have a general negative effect on linguistic performance. Support for this idea comes from the observation that a marginally significant effect is also found for the factor Block (β =-1.273; z = -1.85; p = 0.064). Because slow speech is especially detrimental at the start of the experiment, this suggests that the negative effects of slow speech pertain to task performance in general rather than to performance on particular items.

6.5 Discussion

The experiment investigated whether children's errors in pronoun interpretation are caused by their limited processing speed. The results show that slowed-down speech has a beneficial effect on pronoun comprehension, but only if the child displays the DPBE. This supports the hypothesis that children showing the DPBE do not have sufficient time to take into account the speaker's perspective, causing pronouns to remain ambiguous. The results of the children who already perform correctly on pronouns suggest that in other cases slowed-down speech has an overall negative effect on performance, making the positive effects of slowed-down speech in the DPBE group even more striking.

7 Study 2: Simulation study

We constructed a computational cognitive model to test whether the mechanism of bidirectional optimization can account for children's behavior in the experiment discussed above. To this end, we combined the DPBE/ACT-R model with a computational model of sentence processing. The resulting model, which we refer to as the Speech Rate model, is able to process incoming sentences on a word-by-word basis. With this model we simulated the performance of a group of child participants on sentences with normal and slowed-down speech rate.

7.1 Sentence processing

Words are presented to the model in a serial fashion, with an interval between the consecutive words that is derived from the speech rate. The same sentences are used as in the experiment described above. Two different speech rates were used: a normal speech rate of 4.0 syllables per second (resulting in an inter-word interval of 0.31 seconds) and a slow speech rate of 2.1 syllables per second (inter-word interval: 0.62 seconds). To simulate the differences among utterances in naturally occurring speech, normally distributed noise (m=0, SD=0.01) is added to each inter-word interval. A typical trial commences as follows. As soon as the model detects an audio-event, it focuses its attention on that sound. A word is then retrieved from declarative memory on the basis of the properties of the perceived stimulus. After retrieving the word, its syntactic category is retrieved (for a more extensive description of how concept and lemma information is represented, see van Maanen & van Rijn, 2007). After these retrievals, the word's lexical information is attached to the syntactic goal category that represents the syntactic structure of the sentence (see R. L. Lewis & Vasishth, 2005). As a complete simulation of parsing is not required for investigating the effects of speech rate on the DPBE, this part of the process is implemented in a similar fashion as in the model of reading and dictating of Salvucci and Taatgen (2008).

As soon as the model identifies, on the basis of the retrieved syntactic category, the current word as a pronoun or a reflexive, the model starts the optimization process described earlier (see also Figure 5). So the model does not wait with the process of bidirectional optimization until the sentence is completed but starts the process of bidirectional optimization immediately when it encounters a pronoun or reflexive.

7.2 Selecting the response

After the sentence is processed, the model has to decide whether the sentence is a correct description of the picture. The interpretation of the picture is given to the model from the onset of the trial, as it was also available on the screen before the participants in the experiment heard the sentence. Therefore, the response of the model depends on the outcome of the optimization process, which is the model's interpretation of the anaphor. If the optimization process results in a single interpretation, the model uses that interpretation in its response. However, if the model cannot settle on a single interpretation, it will randomly select a response (with a 80/20 yes-no distribution to reflect the yes-bias, cf. Chien & Wexler, 1990).

Note that this random selection process only takes place when the model cannot settle on an interpretation, that is, when the input is a pronoun and no bidirectional optimization took place. Because the successful use of bidirectional optimization will increase with time, the effect of the yes-bias will gradually decrease. After selecting an answer, the model generates a response by pressing the appropriate button.

7.3 Modeling the acquisition of bidirectional optimization

Because we assume that bidirectional optimization is in principle available, the model develops the ability to perform this process by mere exposure to sentences with pronouns or reflexives. Hereto, we presented the model with randomly selected sentences containing either a pronoun or a reflexive. By means of production compilation, over time the model learns to perform the required operations quicker and with fewer errors.

To simulate the differences in frequency between pronouns and reflexives in natural language, the model was presented with pronouns in 90% of the training trials and reflexives in the remaining 10%. The model was given about 0.32 seconds to determine the optimal meaning for the input, comparable to the time frame in normal speech. As in earlier work on developmental modeling (e.g., McClelland, 1995; van Rijn, van Someren, & van der Maas, 2003), the model was presented with experimental sessions at regular intervals (every 50 trials) to assess the current stage of development. This way, each simulation resulted in 13 simulated experimental datasets. During the "experimental sessions", learning was turned off. This testing scheme was chosen to prevent too much influence of the repeated presentation of the experimental sentences on the outcome.

7.4 Performance of the model on the experiment

For assessing the performance of the model, we ran the model for 16 simulations, resulting in 208 simulated datasets. This way, the effect of speech rate is compared over different simulated participants, who received different amounts of training, thus making the dataset comparable to the human dataset discussed earlier. The same criteria were used to classify the simulated participants into different groups. Of the simulated participants, 97 showed the DPBE (mean number of training trials 177, SD=156) and 110 showed correct performance (mean number of training trials 408, SD=149)[6]. None of the simulated participants showed similar behavior as the children in the Incorrect Performance group, because the model is already able to perform unidirectional optimization from the start.

Similar to the analysis of the experimental data, we analyzed the performance of the simulated participants who showed the DPBE by fitting separate mixed-effect models on the performance on pronoun and reflexive comprehension. The first model contains a random variable to account for the effects of the different simulated participants, and *ExpectedAnswer* (yes or no) to account for the introduced yes-bias. The second model contains the same variables, but also contains the variable *SpeechRate*. A significant difference was found between the mixed-effect models of pronoun comprehension ($\chi^2(2) = 47.801$, $p < 0.001$), but no difference was found between the models of reflexive comprehension ($\chi^2(2) = 0$). Thus,

slowed-down speech has a similar effect on comprehension for the model as for the participants of the experiment. A follow-up analysis on the models' performance on pronoun sentences showed that slowed-down speech did have a beneficial effect on both match (paired t(96)=2.672, p=0.009) and mismatch trials (paired t(96)=5.010, p<0.001).

7.4.1 Model fit

Figure 8 shows the fit of the model with the experimental data of the DPBE group on pronoun sentences (Pearson r^2= 0.96, RMSSD= 1.74).

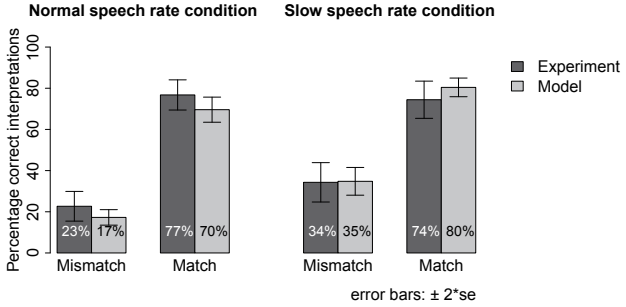


Figure 8: Comparison between mean percentages of correct interpretations of sentences containing a pronoun in the two speech rate conditions, for children (Experiment: n=43) and simulated participants (Model: n=97) showing the DPBE.

The model accounts for the two general trends earlier discussed: the increase in performance on the mismatch items caused by the slowed-down speech, and the large yes-bias. However, one aspect is not captured. The model predicts a significant increase in performance under slowed-down speech for mismatch *and* match items, whereas the experimental data did not show an increase in performance in the match trials. This might suggest that our implementation of the yes-bias is not sensitive enough to capture the details of the child data.

The model's fit on reflexives is not as good as the fit on pronouns. Figure 9 shows the mean percentage of correct interpretations on reflexive comprehension for the model and the experimental data. The model correctly predicts the overall performance on reflexive comprehension for children showing the DPBE. However, the model predicts perfect performance, whereas this level of performance is never found in experimental data with children. Despite these differences, the overall fit (i.e., pronoun and reflexive sentences combined) of the model's performance on the data of the DPBE group is very high (Pearson r^2= 0.96, RMSSD= 2.68).

7.5 Discussion

The computational simulation captures the main effects of the psycholinguistic experiment with children, such as the difference in performance on pronouns versus reflexives, the yes-bias, and the beneficial effect of slowed-down speech on pronoun comprehension in the DPBE

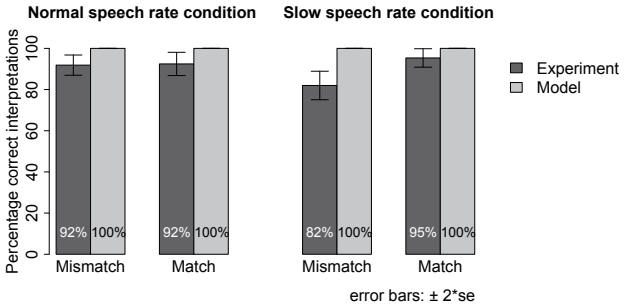


Figure 9: Comparison between mean percentages of correct interpretations of sentences containing a reflexive in the two speech rate conditions, for children (Experiment: n=43) and simulated participants (Model: n=97) showing the DPBE.

group. However, there are also differences between the model and the experimental data. For example, the model predicts perfect performance on reflexive comprehension in the DPBE group. In contrast, the DPBE children in our experiment did not show perfect performance. Another difference between the behavior of the model and children's performance is that the children in our experiment showed a significant decrease in performance with reflexives (DPBE group) and pronouns (Correct Performance group) in the slowed-down speech condition. Children probably have to adjust to the unusually slow speech. The model, on the other hand, is not adjusted to normal speech, and as such does not need to readjust. To summarize, although the model does not explain all details of children's performance in the experiment, it does explain the major effects associated with the DPBE as well as adult-like performance on pronoun and reflexive comprehension.

8 General discussion

In this paper we showed how a linguistic explanation of the DPBE that is embedded in a cognitive architecture allowed us to generate and test detailed predictions with respect to linguistic performance. According to Hendriks and Spenader's (2006) optimality theoretic account of the DPBE, pronouns are ambiguous and are disambiguated only if hearers not only select the optimal meaning for the pronoun, but also take into account the speaker's perspective. This allows them to block the coreferential meaning. We modeled this process in the cognitive architecture ACT-R (Anderson et al., 2004). Our DPBE/ACT-R model simulates adult pronoun comprehension as a process consisting of two consecutive steps. The first step involves selecting the optimal meaning for the pronoun, and the second step involves checking whether a speaker would have expressed this meaning with the same form. Our DPBE/ACT-R model predicts that performing the two steps consecutively requires more processing time than performing only the first step. If children are given sufficient time to perform both steps within the available time, they are predicted to be able to block the coreferential meaning for the pronoun. We tested this prediction by comparing children's comprehension of pronouns at a

normal speech rate with their comprehension at a slower speech rate. Our finding confirms the predictions of the DPBE/ACT-R model: slowed-down speech has a significant beneficial effect on pronoun comprehension, but only if the child displays the DPBE.

If the DPBE were caused by children's lack of pragmatic knowledge, as Thornton and Wexler (1999) argue, it remains unexplained how slowing down the speech rate would provide children with the necessary pragmatic knowledge or the ability to use this knowledge to interpret pronouns correctly. Although Reinhart's (2006) explanation of the DPBE in terms of children's limited working memory capacity appears to be related to the explanation presented here, it is unclear how exactly working memory limitations influence children's comprehension, and how this relates to the present findings. It has been argued that slowed-down speech places a greater temporal load on working memory, because information must be retained over a longer duration (e.g., Small, Andersen, & Kempler, 1997). If this is true, then slowed-down speech is expected to decrease performance when working memory capacity is limited, in contrast to what Reinhart predicts. However, the results of studies investigating the relation between slowed-down speech and working memory are not very clear. For example, Montgomery (2004) did not find an association between sentence processing at different speech rates and working memory capacity in children. So although Thornton and Wexler (1999) and Reinhart (2006) attribute the DPBE to non-linguistic factors, it is difficult to see how these accounts would explain the present findings. In addition, it remains unclear how these accounts relate to general constraints on cognition, and what predictions they would and would not generate regarding children's and adults' linguistic performance.

The results of the psycholinguistic experiment are predicted by the DPBE/ACT-R model, which was constructed by embedding the optimality theoretic account of Hendriks and Spenader (2006) in the cognitive architecture ACT-R. However, these results do not necessarily follow from the optimality theoretic account in itself. Hendriks and Spenader's optimality theoretic account would also be compatible with an explanation in terms of perspective taking: Children may be unable to use bidirectional optimization because they lack the cognitive ability to take into account another person's perspective. In contrast to the explanation implemented in the DPBE/ACT-R model, this explanation would not predict an effect of slowed-down speech, because it is unclear how slowed-down speech would improve children's cognitive skills. Another conceivable explanation of the DPBE that is compatible with an optimality theoretic account is the view that bidirectional optimization is an off-line pragmatic decision process. This view contrasts with our DPBE/ACT-R model, as we implemented bidirectional optimization as a process that takes place *during* on-line sentence processing. If bidirectional optimization is only performed after completion of the sentence, slowed-down speech is not expected to have any effect on comprehension. In the two speech rate conditions, the same amount of processing time was available at the end of the sentence: Participants in the psycholinguistic experiment could take as much time as needed to give a response in either condition. However, within the sentence, processing time was limited due to the presentation of the next word of the sentence, as the critical word (i.e., a pronoun or a reflexive) was always followed by further sentence material in the form of a prepositional phrase. Therefore, the results of this study suggest that the process of bidirectional optimization is an on-line

process.

In addition to a psycholinguistic study, we also performed a simulation study to investigate the predictions of the DPBE/ACT-R model. We built a new cognitive model that also allowed for incremental sentence processing. This model was shown to capture the main effects of slowed-down speech on comprehension that were seen in the psycholinguistic study. For those simulated participants who displayed the DPBE, the cognitive model showed an increase in performance due to slowed-down speech on the comprehension of pronouns, but no effect of slowed-down speech on the comprehension of reflexives. These results support the hypothesis that difficulties with pronoun comprehension are caused by a limited speed of processing, due to which the process of bidirectional optimization cannot be completed.

In our simulations, the process of bidirectional optimization gradually became more efficient as the number of trainings items increased, because the production compilation mechanism of ACT-R is dependent on frequency of use. As a consequence, the model predicts that repetitive testing of children showing the DPBE on pronoun sentences in binding contexts will result in an increase of their performance on pronoun comprehension (although we did not simulate this in our model). However, we assume that children only start to perform bidirectional optimization for pronoun comprehension when their cognitive and linguistic capacities are sufficiently developed (cf. Case, 1987; van Rijn et al., 2003). This is reflected in the starting point of our model, according to which children are in principle able to perform bidirectional optimization, but not yet within the limited amount of time. Therefore, we predict that children will only show a positive effect of repetitive testing and slowed-down speech when they are ready to master the process of bidirectional optimization.

Our simulation study also illustrates some of the considerations and limitations in using cognitive models to study theories of language acquisition. First, cognitive models necessarily are simplifications of reality. Therefore, choices have to be made as to what aspects of the task should be modeled and what aspects can be left unspecified. For example, we chose not to model the sentence-processing component of the model in detail. One of the effects of this choice was that the performance of the model only increased significantly on the comprehension of pronouns at half the normal speech rate. In contrast, the DPBE children showed an increase in performance already at two-third of the normal speech rate. This difference is caused by a simplification of the sentence processing component: In the current version of the model, processing a word takes almost all the time that is available before the next word comes in (about 300 of the 320 ms). Hence, not much time is left for bidirectional optimization. To obtain a significant effect of bidirectional optimization, we had to slow down the speech rate more. However, this simplification of the cognitive model did not result in a qualitative difference between the simulation model and the psycholinguistic study, but only in a quantitative difference. It is left for further study whether a more realistic sentence-processing component would lead to better predictions by the cognitive model.

A difficulty in using cognitive models to study language is the possibility that the linguistic theory and the cognitive architecture may employ different or even conflicting assumptions. For example, Optimality Theory, due to its roots in neural network theory, assumes candidates to be evaluated in parallel, and also assumes the constraints of the grammar to apply

in parallel. ACT-R, on the other hand, assumes a central processing bottleneck. This implies that only one production rule can be applied at a time. We chose to adopt the ACT-R assumption, since it imposes the strongest restrictions on cognitive processing. Note that this choice is not incompatible with OT per se, as it preserves the input-output relations predicted by the OT grammar as well as the linguistic knowledge constraining these relations, but merely specifies the process by which these input-output relations are obtained. As a result of this choice, in the DPBE/ACT-R model only two candidates are evaluated at a time and the constraints are applied one by one. The hypothesis that children do not have sufficient time to perform bidirectional optimization follows from this particular property of the DPBE/ACT-R model.

A related issue concerns those cases where a particular effect could in principle be explained by the grammar, but also by the cognitive architecture. In language acquisition research, computational models of grammar typically use corpus data as input and observed patterns in the child's speech as output. As a consequence, frequency distribution patterns in the input and output are of crucial importance to the grammar. In a cognitive modeling approach, the grammar may be non-probabilistic because the cognitive model already is sensitive to frequency distributions. For example, our DPBE/ACT-R model was trained on language input which consisted of 10% reflexives and 90% pronouns, reflecting the unequal distribution of reflexives versus pronouns found in corpus studies of child-directed speech (e.g., Bloom, Barss, Nicol, & Conway, 1994). Because the production compilation mechanism of ACT-R is dependent on frequency of use, this unequal frequency distribution resulted in a faster acquisition of bidirectional optimization for pronouns than for reflexives (although the model assigns a correct interpretation to reflexives faster, because its interpretation is not dependent on bidirectional optimization). So cognitive modeling accounts of language acquisition are not incompatible with frequency-based accounts, but rather provide complementary insights. The exact division of labor between grammar and cognitive architecture may be determined by theory-internal considerations as well as empirical observations.

In conclusion, embedding a theory of linguistic competence in a cognitive architecture is a promising new approach to understanding issues in the domain of language. While linguistic theories may offer an adequate account of children's linguistic competence, cognitively informed models are required to test these competence theories empirically. Because cognitive architectures are based on well-founded theories of cognition and guide the construction of computational simulations that allow us to test the performance of a cognitive system under different conditions, they may help us to gain a better understanding of the process of language acquisition.

Notes

[1] The definition of c-command used here is: Node A c-commands node B if the first branching node of the syntax tree that dominates A, also dominates B.

[2] A $\gg$ B means that constraint A is higher ranked, i.e., stronger, than constraint B.

[3] For an overview of existing ACT-R models, see: http://actr.psy.cmu.edu/publications

[4] The code of the discussed ACT-R models can be retrieved from: http://www.let.rug.nl/jacolienvanrij/modelcode.html

[5] In earlier papers (van Rij, Hendriks, Spenader, & van Rijn, 2009*a*, 2009*b*), we distinguished four different groups: i) the No DPBE group (n=5), ii) the Extra-Linguistic Strategy group (n=9), iii) the DPBE group (n=34), and iv) the Correct Performance group (n=14). Participants who were classified as belonging to the Extra-Linguistic Strategy group used the extra-linguistic strategy of answering 'yes' to all pronoun mismatch items in both speech rate conditions, while their performance on reflexive items was correct. Participants who were classified as belonging to the DPBE group did not seem to make use of a particular strategy for answering the pronoun items, sometimes giving a correct response while at other times giving an incorrect response, although they showed a general bias to say 'yes'. For simplicity, we combined the Extra-Linguistic Strategy group with the DPBE group in this paper. In addition, we changed the name of the No DPBE group into Incorrect Performance group, because this name better reflects the behavior of its members.

[6] Only 207 simulated participants are reported (97 DPBE and 110 Correct Performance), because the final experiment of one of the simulations was interrupted, resulting in only 12 experimental datasets for that particular simulation.

chapter 3

A cognitive model of the acquisition and use of referring expressions:

Effects of working memory and perspective taking

This manuscript was submitted as:
Van Rij, J., Van Rijn, H., & Hendriks, P. (submitted). A cognitive model of the acquisition and use of referring expressions: Effects of working memory and perspective taking. Manuscript, October 2011.

Abstract

In this paper we present a cognitive model of the acquisition and use of referring expressions in discourse. In certain discourse contexts, 4- to 7-year-old children produce unrecoverable pronouns where a full noun phrase would have been the adult choice. Also in their comprehension of pronouns in discourse, these children make non-adult choices. Based on computational simulations, we argue that mature use of referring subjects requires (a) sufficient working memory capacity to identify the discourse topic, and (b) sufficient speed of processing to determine whether the intended meaning is recoverable for the listener from the expression used. The cognitive model allows us to investigate the complex interaction between linguistic constraints and cognitive factors. In addition, the model generates detailed and testable predictions with respect to further linguistic performance.

1 Introduction

When speakers wish to refer to a character, event or object that has already been introduced in the context, they can choose between different referring expressions. For example, they can use a proper name, a definite noun phrase or a pronoun when referring to a character. These expressions differ, on the one hand, in the type and amount of information they convey, and, on the other hand, in their structural complexity and length. How do speakers determine which referring expression fits a particular situation best? Consider the following story:

1. A soccer player misses a penalty in the final.
2. He agrees to give an interview after the match.
3. A reporter asks the soccer player some sharp questions.

4a. He thinks the team didn't play well.

4b. The soccer player thinks the team didn't play well.

Figure 1: Example story with two alternatives for the last sentence.

In Sentence 2, a subject pronoun (*he*) is generally preferred over a definite noun phrase (*the soccer player*) for reference to the soccer player. Contrastively, the unstressed subject pronoun in Sentence 4a cannot be used for reference to the soccer player because a pronoun in this position will be interpreted as reference to the reporter. Rather, a definite noun phrase must be used for reference to the soccer player, as in Sentence 4b. What processes do speakers use that makes them prefer a pronoun in Sentence 2, but a definite noun phrase in Sentence 4? And if these processes have to be learned, can we find signatures of this development in children?

Two influential linguistic approaches have been put forward to account for the choice of referring expressions. The first approach, which we will refer to as the *discourse-oriented approach*, assumes that the choice of referring expression is crucially dependent on the properties of the linguistic discourse (a.o., Ariel, 1988; Givón, 1983). More specifically, the saliency or *accessibility* of a character in the discourse determines which referring expression is appropriate to use. For example, pronouns are used for reference to the character that is most accessible in the current discourse, and full noun phrases are used for characters that are less accessible. So according to the discourse-oriented approach, a speaker selects a referring expression based on the accessibility of the referent. A listener, in turn, can infer the intended referent on the basis of the implied accessibility level and the mutual knowledge of speaker and listener. For Sentence 2 this means that a speaker will use a pronoun for reference to the soccer player because at this point the soccer player is the only and thus most accessible referent in the discourse. For Sentence 4, however, a pronoun cannot be used for reference to the soccer player because now the reporter is the most accessible referent. The second approach, the *listener-oriented approach*, provides an alternative explanation for the choice of referring expressions (a.o., Gundel et al., 1993; Hendriks et al., 2008). In contrast to discourse-oriented accounts that assume a focus on the speaker's mental states, this approach assumes that when

more than one referring expression can be used, a speaker takes into account the listener's perspective. In particular, Gundel et al. argue that speakers apply Grice's Maxim of Quantity (Make your contribution as informative as required, but do not make your contribution more informative than is required, cf. Grice, 1975). According to this approach, the speaker cannot use a pronoun for reference to the soccer player in Sentence 4 because a *listener* would interpret this pronoun as referring to the reporter as the reporter is the most salient referent.

In sum, the two approaches propose different mechanisms to explain the same phenomena. It is difficult to exclude one of these approaches in favor of the other on the basis of linguistic considerations only because the accounts often make the same predictions. However, the two approaches differ in the way selection of the appropriate referring expression is related to cognitive factors. The discourse-oriented approach associates the accessibility of the discourse referents with the accessibility of representations in memory, so that referents with lower accessibility take more time to retrieve (Ariel, 1988). In this regard, this approach is a linguistic variant of sentence-processing theories that explain language processing in terms of memory processes and activation of representations (a.o., Just & Carpenter, 1992; R. L. Lewis & Vasishth, 2005; Sprenger, Levelt, & Kempen, 2006). The listener-oriented approach, on the other hand, does not make specific predictions about processing, but assumes that taking into account the listener's perspective requires Theory of Mind-like abilities (Gundel, Ntiletheos, & Kowalski, 2007). As a consequence of these differences, the two linguistic approaches make different predictions about the acquisition of referential choice, to which we will return later.

In this paper we show that the two approaches are not necessarily incompatible nor can one be reduced to the other. Rather, we argue that both accessibility and perspective taking are necessary to account for the choice of referring expressions. Within the framework of ACT-R (Anderson et al., 2004), we developed a computational cognitive model that simulates the production and interpretation of referring expressions in subject position, expanding on and complementing our earlier model of referring expressions in object position (van Rij, van Rijn, & Hendriks, 2010). Our computational model integrates the main ideas of the discourse-oriented account and the listener-oriented account, as illustrated in Figure 2.

In our computational model, the accessibility of the discourse referents is modeled as activation of representations in memory. The level of activation is determined by the preceding discourse. To construe an adult-like representation of the discourse, sufficient working memory capacity is required. After selecting an initial referring expression to use, the model checks whether this referring expression will be interpretable for the listener. If not, the referring expression is discarded and another referring expression is selected. The speaker's choice of referring expression is thus implemented as a Theory of Mind-like process that is performed during on-line production. Crucially, this process requires sufficient processing speed, which is obtained through linguistic experience (van Rij et al., 2010). Thus, our model associates the accessibility of the discourse referents with the accessibility of representations in memory, as in the discourse-oriented approach, and uses this accessibility to evaluate whether a referring expression is interpretable for the listener by taking into account the listener's perspective, as in the listener-oriented approach.

By integrating ideas from both approaches, this cognitive model provides a novel expla-

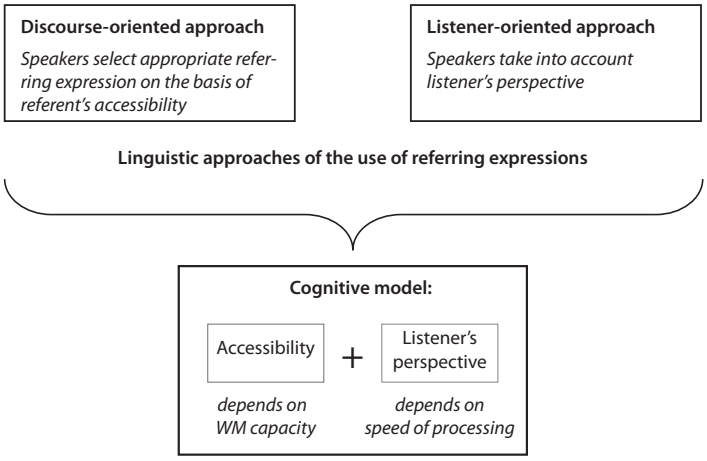


Figure 2: In our cognitive model we combine two influential linguistic accounts to explain the use and interpretation of referring expressions, and their acquisition.

nation of how adult speakers determine what is the best referring expression to use in a particular context. Importantly, we can formulate precise and testable predictions with respect to the use and acquisition of referring subjects in discourse, by computationally simulating how the processes of reference production and comprehension are learned. On the basis of simulations that are compared to empirical data of 4- to 7-year old children and adults, we argue that sufficient WM capacity as well as sufficient processing speed is required for adult-like performance on the production and comprehension of referring subjects.

This paper is organized as follows. We start with a presentation of the main empirical phenomena related to referring subjects. Subsequently, we present our cognitive model and describe the modeling decisions we made. Finally, we discuss the main results of the simulations run with our model, and the implications of this work for our understanding of the use of referring expressions.

2 Empirical phenomena

In this section we briefly discuss relevant experimental results of adults' and children's use and interpretation of referring expressions, and suggest possible explanations for children's difficulties with referring expressions.

2.1 Adults' use and interpretation of referring expressions in discourse

In general, speakers use a pronoun for reference to highly prominent referents in the discourse and use more specific forms, such as a full noun phrase (NP), for less prominent referents (e.g., Ariel, 1990; Grosz, Weinstein, & Joshi, 1995; Gundel et al., 1993). Previous research

has identified various discourse factors that affect the use and interpretation of referring expressions, including how recently a referent was used (e.g., Arnold, 1998; Clark & Sengul, 1979; Ehrlich & Rayner, 1983; Givón, 1983), the grammatical role of the referent (e.g., Arnold, 1998; Brennan, 1995; Crawley, Stevenson, & Kleinman, 1990; Grosz et al., 1995; Kaiser & Trueswell, 2008), the order of mentioning of the referents (e.g., Gernsbacher & Hargreaves, 1988; Gordon, Grosz, & Gilliom, 1993; Kaiser & Trueswell, 2008), verb semantics (e.g., McDonald & MacWhinney, 1995), and parallelism of the syntactic roles of pronouns and their potential antecedents (e.g., Chambers & Smyth, 1998; Grober, Beardsley, & Caramazza, 1978). In addition to the characteristics of the linguistic discourse, experimental results suggest that speakers try to avoid expressions that would result in interpretation difficulties for the listener (e.g., Arnold & Griffin, 2007; Brennan & Hanna, 2009; Clark, 1996; Gundel et al., 1993; Haywood, Pickering, & Branigan, 2005; Heller, Grodner, & Tanenhaus, 2008; Hendriks et al., 2008; Tily & Piantadosi, 2009). However, the speaker's estimation of the listener's interpretation seem to have limitations because speakers do not avoid ambiguities in every situation (e.g., Bard et al., 2000; Ferreira & Dell, 2000; Horton & Keysar, 1996; Keysar, Barr, & Horton, 1998), and sometimes use overspecified expressions when there is no ambiguity for the listener (e.g. Arnold & Griffin, 2007).

So, where discourse and perspective taking both seem to play a role in adult's production and comprehension of pronouns, children do not use discourse information and perspective taking in the same way as adults do. The differences between children's and adults' performance may provide insight into the cognitive processes that play a role in the use and acquisition of referring expressions.

2.2 Children's production of referring expressions in discourse

Children have been found to make errors with the production of referring expressions up to a relatively late age, sometimes as late as age 6. Using story-books with pictures to elicit narratives, Karmiloff-Smith (1985) found that English and French speaking children younger than seven used mainly pronouns, even if the use of a pronoun may be interpreted incorrectly by a listener who does not have access to the pictures. The question is what causes this overuse of pronouns.

Children are able to use discourse factors such as the accessibility of referents when choosing a referring expression. Hickmann and Hendriks (1999) found that children of 6, 9 and 11-year old used significantly more pronouns for a referent that was prominent in the story than for the other referents in the story. However, a recent study by Wubs et al. (2009) shows that children have difficulties when the listener's perspective has to be taken into account when choosing a referring expression. In this study, children aged 4-6 and adults were asked to tell stories about two characters on the basis of series of pictures that were designed to elicit two topic shifts. In the first few pictures of each story, only one of the two characters is visible. The first topic shift occurs halfway through the story, after the introduction of the second character, and results in the second character becoming the new discourse topic. The second topic shift, back to the initial character, is initiated at the end of the story as the

final picture focuses on the first character again. Crucially, as both referents are of the same gender, using a pronoun to refer to the first character while the second character is the discourse topic will result in potential ambiguity. Wubs et al. report that children mostly used pronouns for reference to the character that is not the discourse topic, which would result in an incorrect interpretation for the listener. To avoid an incorrect interpretation for the listener, adult participants only used full NPs. Wubs et al. concluded on the basis of these results that children fail to take into account the listener's perspective in language production, resulting in an over-use of pronouns.

2.3 Children's interpretation of referring expressions in discourse

When looking at children's performance in the comprehension of referring expressions, a more mixed picture emerges. Some studies find that children are sensitive to discourse factors from a very young age, whereas other studies report that children have difficulties incorporating discourse factors. Song and Fisher (2005, 2007) argue that even very young children use similar factors as adults to determine the saliency of referents in the discourse. They report that their 2.5 and 3-year old participants showed a preference for looking at the most salient referent a second after having heard an ambiguous subject pronoun at the end of short stories. However, Arnold, Brown-Schmidt, and Trueswell (2007) compared adults' and 3 to 5 year old children's processing of subject pronouns following a sentence that introduced two referents. When the pronoun was ambiguous, adults showed a clear preference for the first-mentioned referent as the antecedent of the ambiguous pronoun. Children, however, did not show a first-mention bias in their answers and in their gaze behavior (Arnold et al., 2007). Wubs et al. (2009) also investigated how discourse structure influences the interpretation of pronouns in short stories. The children and adults that participated in the earlier mentioned production task also performed a comprehension task in which the participants listened to pre-recorded stories involving two characters. The stories differed in whether they did or did not contain a topic shift. The final sentence of each story contained a potentially ambiguous subject pronoun, followed by a comprehension question of which the answer is dependent on the interpretation of that pronoun. Wubs et al. (2009) found that in most stories with a topic shift, adults selected the new discourse topic (i.e., the second character) as the antecedent for the pronoun, and in almost all stories without a topic shift they selected the initial discourse topic (i.e., the first character). In contrast, children did not use the prior discourse structure, as their performance was not influenced by the presence or absence of a topic shift. Thus, children do not use the linguistic information that signals a topic shift in the same way as adults do.

To summarize, although children from a very young age display sensitivity to some discourse factors (e.g., Hickmann & Hendriks, 1999; Song & Fisher, 2005, 2007), children up to age 6 show difficulties with pronoun use and interpretation (Arnold & Griffin, 2007; Wubs et al., 2009).

2.4 Explanations for children's non-adult performance

An important question is whether children's deviations from adult-like performance in comprehension and production can be explained by a single set of mechanisms (Karmiloff-Smith, 1985; Kempen, Olsthoorn, & Sprenger, 2011; Wubs et al., 2009). As mentioned earlier, if the accessibility of discourse referents reflects the activation of memory representations (e.g., Ariel, 1988), working memory limitations may give rise to non-adult performance on pronoun production and comprehension. On the other hand, it has been proposed that Theory of Mind skills are required for adult-like performance on pronoun production, as speakers need to take into account the listener's perspective (Gundel et al., 2007). On the basis of empirical findings we argue that both insufficient working memory capacity and underdeveloped Theory-of Mind skills may cause non-adult performance.

2.4.1 Working memory

To investigate whether comprehension and production of referring subjects are affected by processing limitations. Wubs et al. (2009) also measured children's and adults' working memory (WM) capacity. Koster, Hoeks, and Hendriks (2011) report that children with higher WM scores performed more adult-like in the production and comprehension tasks. Similarly, Hendriks et al. (2008) found a correlation between a performance measure of the production of referring subjects, and scores on the working memory (WM) task in elderly participants: the lower the WM score, the more pronouns were produced in situations where a full NP would have been more appropriate. These results suggest that adult-like production and comprehension of referring subjects requires sufficient WM capacity. Moreover, the performance of young adults was not affected by WM (Koster et al., 2011) suggesting that performance is not influenced if the WM span is above a certain threshold.

2.4.2 Theory of Mind

The question arises whether, besides WM capacity, also Theory of Mind skills are crucial for adult-like use of referring expression. If speakers must consider the listener's perspective to check whether the intended referring expression is recoverable, they will need mature Theory of Mind skills as they must be able to predict and reason about the beliefs, knowledge, intentions, emotions and desires of other people (Premack & Woodruff, 1978). Children below 4 years old show difficulties in predicting other people's beliefs on false-belief tasks (Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983, but see Onishi & Baillargeon, 2005). Studies that investigate the language skills of children and adults with autism support the hypothesis proposed by Gundel et al. (2007) that Theory of Mind is required for adult-like production of referring expressions. Children and adults with autism, who have difficulties with Theory of Mind reasoning (a.o., Baron-Cohen, Leslie, & Frith, 1985), also have problems in producing a contingent discourse. These problems with discourse production are closely related to their difficulties in performance on Theory of Mind tasks (Hale & Tager-Flusberg, 2005). In addition, Arnold, Bennetto, and Diehl (2009) report that children with autism (ages 9-12) produce

more overspecified referring expressions (i.e., producing a NP where a pronoun would have sufficed) than adolescents with autism (ages 13-17) and control participants of the same age. These differences in performance cannot be explained solely by the lower WM score of participants with autism because the adolescents with autism, who also have lower WM scores, did not produce more overspecified referring expressions than control participants (Arnold et al., 2009). These findings suggest that adult-like use of referring expressions in discourse requires mature Theory of Mind skills. This is in line with the listener-based approach to referential choice, according to which speakers use a Theory-of-Mind-like process to avoid producing referring expressions that are unrecoverable for the listener. Therefore, an adequate account of referential choice should combine the dependency of mature production and comprehension of referring expressions on sufficient WM capacity with the need for the speaker to take into account the listener's perspective.

2.4.3 Processing speed

To investigate how adult-like performance is achieved, we have extended our previous cognitive model, which explained children's difficulties with perspective taking by their limited speed of processing (van Rij et al., 2010). The basic idea of this model is that children's processing is not efficient enough yet to take into account the conversational partner's perspective if the available time is limited, as in on-line comprehension. The results of our previous experimental study with object pronouns (van Rij et al., 2010) support the idea that children's development of the interpretation of pronouns is linked to an increase in their speed of processing: Only when children have acquired sufficient processing speed, they can use perspective taking within the available time and interpret object pronouns in an adult-like way. Whereas the use of object pronouns is strongly influenced by syntactic constraints, the use of subject pronouns is more dependent on the preceding discourse. We have extended our earlier cognitive model so that it explains the acquisition and use of subject pronouns in discourse too and can also account for the WM effects that have been found in various studies (a.o., Hendriks et al., 2008; Wubs et al., 2009). On the basis of our extended computational model, we argue that the speaker needs to determine the current discourse topic and needs to take into account the listener's perspective to decide which referring expression to use. We predict that sufficient WM capacity is necessary for determining the discourse topic. On the other hand, Theory of Mind skills and sufficient processing speed are required for applying perspective taking during on-line production.

3 A computational model of the production and comprehension of referring expressions

In this section we discuss our computational model and the results of the simulations of this model. Before describing the actual implementation and the performance of the model, we will first introduce the cognitive architecture ACT-R, in the context of which our model is implemented.

3.1 ACT-R model

Cognitive models are computational simulations of the cognitive processes involved in performing a task, for example comprehending a sentence (e.g., Budiu & Anderson, 2004). Our model of the acquisition and use of referring expressions is implemented within the cognitive architecture ACT-R (Adaptive Control of Thought-Rational; Anderson et al., 2004). ACT-R's modeling environment constrains simulation models on the basis of built-in and well-tested mechanisms and parameters to ensure psychological plausibility. As a theory, ACT-R has a modular structure: Each of ACT-R's modules is based on smaller theories on cognition. The two main modules of ACT-R are declarative memory and the central production system (a description of the other modules and a more extensive description of the architecture can be found in Anderson et al., 2004).

The declarative memory module contains chunks of factual information. Chunks can be connected with other chunks, forming a network. Each chunk has an activation value that is based on a weighted average of the number of prior occurrences of that chunk. This activation decays with time. However, activation is not only determined by the chunk's own history, but also by the usefulness of that chunk in contexts similar to the current context. Chunks that are retrieved or manipulated spread activation to all chunks they are connected with. So the activation of a chunk is the summation of its own base level activation and the spreading activation from other chunks that are currently being processed. The activation of chunks determines both the probability that this chunk will be retrieved and the time required for retrieval.

In the ACT-R framework, the central production system contains IF-THEN rules. The IF-clause of each production rule specifies the conditions that must be met for that production rule to be considered for execution (for example, the presence of a certain stimulus in the visual system, or a linguistic form that requires an interpretation). The THEN-clause specifies which actions will be performed if that rule is selected for execution (for example, a key press or the retrieval of a memory element). At each time step, the central production system matches the production rules to the current state of the system, and the most active matching rule is selected for execution. The activation value of production rules reflects the utility of that rule and is an expression of the expected benefits of executing that production rule discounted for the costs associated with that production rule.

3.2 Implementation of perspective taking

Our cognitive model simulates the use and interpretation of referring expressions. We have implemented a linguistic account of the use of referring subjects in terms of optimization described below in ACT-R as an extension of our cognitive model of the acquisition and use of object pronouns (van Rij et al., 2010). This account explains children's non-adult-like performance as resulting from a constraint-based grammar in combination with children's inability to take into account the opposite perspective in conversation. Figure 3 provides a schematic overview of how the model produces a referring subject.

Choosing a referring expression requires several steps: First, the model needs to deter-

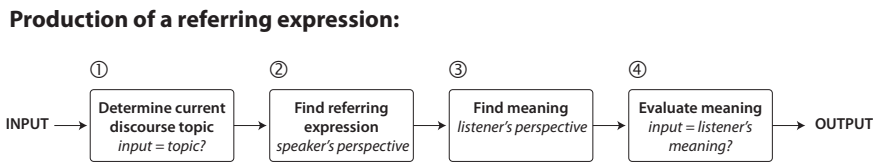


Figure 3: ACT-R model of producing a referring expression (output) for a given referent (input) by combining optimization from the speaker's perspective (step 2) with optimization from the listener's perspective (step 3).

mine what is the current discourse topic on the basis of the prior discourse (Figure 3, step 1). Furthermore, the model needs to apply perspective taking to determine which referring expression to use (Figure 3, Step 2-4). As a speaker, our model tries to find the optimal form (either a pronoun or a full NP) for reference to an input referent that either is or is not the discourse topic (Step 2). To determine whether the listener can correctly interpret this expression, in a next step (Step 3) the model takes the perspective of a listener. As a listener, the model optimizes in the opposite direction, starting from the selected expression (a pronoun or full NP) and deriving the optimal interpretation of that form (i.e., reference to the discourse topic or to a non-topical referent). The model then evaluates whether the selected referring expression leads to the intended interpretation for the listener and decides whether it should use that expression or discard it (Figure 3, Step 4). In this section, we will first discuss how the model learns to consider the opposite perspective in communication (Step 3-4). Following this discussion, we will explain how the model determines the current discourse topic (Figure 3, Step 1).

In Step 2 in Figure 3, our model selects the optimal referring expression or meaning from a set of candidates that best satisfies the ranked and violable constraints of the language (cf. Optimality Theory; Prince & Smolensky, 2004). This selection process makes use of two linguistic constraints that have been proposed to guide the production and comprehension of pronouns in discourse: ProTop and Referential Economy. ProTop is a constraint stating that pronouns refer to the discourse topic (Beaver, 2004; Grosz et al., 1995; Hendriks et al., 2008). As it establishes an association between a particular form (a pronoun) and a particular meaning (reference to the topic), this constraint has an effect in both production and comprehension. In production, however, the stronger constraint Referential Economy also plays a role. Referential Economy (cf. Burzio, 1998; Hendriks & Spenader, 2006) expresses the idea that it is more economical to use a pronoun than a full NP, regardless of whether the referent is the topic or not. As this constraint expresses a preference among forms irrespective of their meanings, this constraint only has an effect in production. Since Referential Economy is stronger than ProTop, the combination of these two constraints will result in a general preference for pronouns over full NPs in production.

Figure 4 shows how the model uses linguistic knowledge when selecting the referring expression to be used (Step 2 in Figure 3) or when selecting the best interpretation for a certain

referring expression (Step 3 in Figure 3).

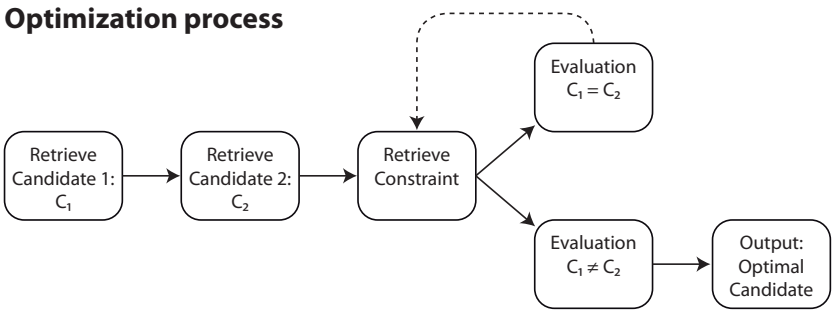


Figure 4: Structure of the ACT-R model of linguistic optimization. The same optimization process is used for production (Step 2 in Figure 3) and comprehension (Step 3 in Figure 3). From a set of candidate outputs, production optimization (i.e., optimization from the speaker's perspective) selects the optimal form and comprehension optimization (i.e., optimization from the listener's perspective) selects the optimal meaning.

In our model, candidate forms, candidate referents, and linguistic constraints are all represented as chunks in declarative memory. For any given input, the model first retrieves two possible candidates from declarative memory, followed by the retrieval of a relevant constraint of the grammar (see Figure 4). The system will retrieve the highest ranked constraint first because the constraint ranking reflects the activation of the chunks. On the basis of the retrieved constraint the model evaluates the two candidates. If one of the candidates violates the constraint but the other candidate does not, the model selects the other candidate as the optimal candidate. If both candidates either violate or satisfy the constraint and hence the constraint does not distinguish the candidates, then the model retrieves the next ranked constraint to evaluate the candidates. If the model cannot retrieve another relevant constraint, for example because there are no more relevant constraints left or because time is up, one of the two candidates is selected at random as the optimal candidate. The model uses a greedy optimization process by only evaluating two candidates at a time until one of the two evaluated candidates is preferred over the other. This way, the model needs less time (compared to an exhaustive search) to find a solution, although the solution that is found might be suboptimal. If more time is available, the model will try to find a better solution by comparing the optimal candidate at that point with additional candidates. Note that this approach hinges on the assumption that the previous discourse increases the activation of the relevant candidates, since the candidates with the highest activation will be evaluated first.

In our model, perspective taking consists of at least two steps of unidirectional optimization in opposite directions (i.e., *bidirectional optimization*), which are performed serially during on-line sentence production and comprehension (cf. van Rij et al., 2010). In production,

the model starts with the speaker's perspective. In comprehension, the model initially takes the listener's perspective. The output of this initial process serves as the input for another process, in which the model takes the opposite perspective (see Figure 3). For example, if the model's goal is to *produce* the optimal expression for referring to a particular referent, the model starts out from the perspective of a speaker. Optimization proceeds in the model by first retrieving two candidates, a pronoun and a full NP, and a constraint, REFERENTIAL ECONOMY. On the basis of REFERENTIAL ECONOMY, the model determines that a pronoun is the optimal form. This form is the input to another step of optimization (Step 3 in Figure 3), in which the model takes the perspective of the listener and checks whether this form, a pronoun, is recoverable for the listener. From the listener's perspective, the constraint PRO-TOP determines that a pronoun will be interpreted as referring to the discourse topic. As the input of the initial step and the output of this step (Step 3 in Figure 3) are identical, the model concludes that the meaning of the pronoun can be recovered by the listener.

In particular cases, the meaning may not be recoverable for the listener from the selected form even after the second step of perspective taking. In such cases, if time permits, the model considers whether the meaning is recoverable for a *bidirectionally* optimizing listener by performing a third step of optimization. This more complex process of perspective taking can be compared with second-order Theory of Mind reasoning. If the selected form is also not recoverable for a bidirectionally optimizing listener, the model starts a new process of optimization, while blocking the unrecoverable form from being a possible candidate.[1]

Because bidirectional optimization consists of two or more steps of unidirectional optimization, bidirectional optimization takes more time to complete than unidirectional optimization. In comprehension, the amount of time available for selecting an interpretation is limited because the speaker determines the speaking rate. In production, more time may be available for selecting a form, but still the amount of time is limited because of the expectancy of reasonably fluent speech. An important assumption of our model is that, initially, the model can usually only perform a single step of unidirectional optimization within the amount of time available.[2] As a result, the model as a speaker only produces pronouns as referring subjects, and as a listener the model fails to interpret the use of a full NP as indicating a topic shift. However, the ACT-R architecture assumes a learning mechanism called production compilation (Taatgen & Anderson, 2002) that makes cognitive processing more efficient as a result of experience. Production compilation combines two production rules that are repeatedly executed in sequence into one new production rule. This new production rule takes less time to execute than the two old rules. Therefore, the model will learn to prefer the new rule over the two old rules. This new production rule can again be combined with other productions rules, so that eventually the processes of comprehension and production become so efficient that bidirectional optimization can be completed within the limited time (cf. van Rij et al., 2010). In other words, because of this architectural learning mechanism, processing speed increases and therefore performance increases as a function of linguistic experience.

This linguistically motivated computational model of grammar determines optimal forms and meanings on the basis of the constraints of the grammar and the discourse status of potential referents. Although the grammar uses constraints referring to discourse status, such

as PROTOP, the grammar does not determine which referent in the discourse is the topic. This selection procedure is modeled as a separate process, which is explained in the next section.

3.3 Implementation of discourse processing

To be able to computationally simulate children's and adults' production and comprehension of referring expressions, the model must know whether the referent is the discourse topic (see Figure 3, step 1). The discourse topic is generally considered to be what the linguistic context is about (a.o., Ariel, 1990; Arnold, 1998; Givón, 1983; Grosz et al., 1995; Gundel et al., 1993). In our model, the topic is simply the discourse referent with the highest accessibility. The accessibility of a discourse referent is determined by the activation of the chunk in declarative memory representing the discourse referent (cf. Arnold, 2008; Foraker & McElree, 2007). This activation is dependent on the structure of the discourse.

Our model creates a representation of the discourse while it processes the sentence. First, the model incrementally builds up the syntactic structure of the sentence. To this end, sentence processing is implemented as an incremental process of very efficient memory retrievals (cf. R. L. Lewis & Vasishth, 2005). Every time a word is encountered, lexical and syntactic information about the incoming word is retrieved and stored as chunks in declarative memory. On the basis of this information, the model retrieves a syntactic structure to which the word can be attached. For example, if the incoming word is the noun *soccer player* following the determiner *the*, the model retrieves an earlier created syntactic chunk representing a noun phrase, with *the* as the determiner and an empty position for the noun. The encountered word is then attached to the retrieved syntactic chunk, and this modified chunk is attached to the syntactic representation of the sentence.

In addition, the model incrementally builds up the discourse structure using the lexical concepts accessed by the words in the sentence. When the model retrieves a lexical concept for a noun for the first time, the model creates a new discourse referent. This discourse referent contains a link to the lexical concept and includes features such as animacy, gender, and number. The activation value of a discourse referent is determined by the recency of the last retrieval, the history of retrievals, and the association with other chunks. When a lexical concept is retrieved for which a discourse referent already exists, the model does not create a new discourse referent but re-activates the existing chunk. Thus the representation of the discourse is not only based on the prior sentence but on all sentences in the current discourse (in contrast to, for example, Grosz et al., 1995). However, the activation of a discourse referent decays over time. Therefore, the probability to retrieve a referent that is introduced many sentences earlier decreases, unless that chunk is re-activated again.

By modeling accessibility as activation, the accessibility of discourse referents in our model is determined by cognitive rather than linguistic constraints. The grammar uses this gradient information to make all-or-none decisions about candidate forms and meanings through the notion of discourse topic.

The discourse status of referents plays a different role in production and comprehension.

In production, the model first checks whether an input referent is the discourse topic or not by comparing the intended referent with the most active discourse referent. The resulting discourse information (i.e., whether the input meaning is the discourse topic or a non-topical referent) is added to the input of the optimization process. In comprehension, discourse status is used to rank candidate meanings for a given input form. If the model encounters a pronoun, the model first retrieves the discourse topic as a possible referent. Only if the subsequent optimization process reveals that the input expression cannot be interpreted as reference to the topic, will the model retrieve a non-topical referent. In contrast, if the model encounters a full NP, discourse status is not used to determine the referent. The model can retrieve the meaning of the full NP on the basis of the information from declarative memory and use that meaning to access the associated discourse referent.

Since the discourse topic is defined as the most active discourse referent, discourse status is dependent on activation. In our model, the probability of being the discourse topic is high when the referent has been mentioned often or when the referent has been mentioned recently (cf. Arnold, 1998). In addition, we propose that spreading activation from other discourse referents also plays a role and is crucially influenced by WM capacity, as explained below.

3.4 Implementation of working memory capacity

ACT-R assumes no separate WM component (but see Borst, Taatgen, & van Rijn, 2010). However, Daily, Lovett, and Reder (2001) proposed that WM capacity can be explained in ACT-R by the amount of *source activation*. Source activation is the activation spreading from the current goal to other chunks, so that information associated with the goal is more accessible. The goal contains information to keep track of the current state in performing a task. Chunks that are retrieved or manipulated are placed in one of the buffers of the ACT-R system, for example the goal buffer. The production rules can only access and manipulate chunks from the different modules when they are placed in a buffer, which can hold only one chunk at a time. Crucially, only the chunks in the buffers spread activation.

According to Daily et al. (2001), a low WM capacity can be explained by a relatively small amount of spreading activation from the current goal, whereas a high WM capacity can be explained by a large amount of spreading activation. In the domain of language production, Reitter (2008) uses this theory of spreading activation to explain short-term priming of syntactical structures, and R. L. Lewis and Vasishth (2005) assign memory processes a central role in their model of sentence processing. We implemented Daily et al.'s spreading-activation account of WM to explain the observed correlation between children's performance on the use and comprehension of referring expressions and their WM capacity. We propose that differences in participants' WM capacity can be accounted for by differences in spreading activation from the sentential subject. Subjects of a sentence are highly prominent elements, not only within their own sentence but also in the discourse. Pronouns are preferably interpreted as referring to the subject of the previous sentence (a.o., Arnold, 1998; Grosz et al., 1995; McDonald & MacWhinney, 1995). To implement this idea, the model keeps the subject of the

previous sentence active as part of the current goal, as a simplified representation of the previous sentence.

Recall that the activation of chunks is the summation of their own base-level activation, determined by the history of the chunk, and the spreading activation of all other chunks that currently reside in one of the buffers. The subject of the previous sentence is proposed to be an additional source of spreading activation and to spread activation to all discourse referents associated with the subject. If a participant simulated by the model has a high WM capacity, the previous subject spreads a large amount of activation and, as a result, discourse referents that are associated with the subject become more activated. As a consequence, the previous subject is more likely to be selected as the antecedent for the current pronoun. If the subject spreads a small amount of activation, as is the case for a simulation of a low WM capacity participant, then there will be a neglectable effect on the discourse referents associated with the subject. In that case, the effects of frequency and recency will be the main determinants of the activation of the discourse referents.

To summarize, in our cognitive model the activation of a discourse referent is determined by the recency and frequency with which the chunk is mentioned in the discourse, and the spreading activation from the subject of the previous sentence. The model uses the information about the activation of the discourse referents in the process of perspective taking to select a referring expression or to find an interpretation for a given referring expression.

On the basis of our model's simulations, we argue that both sufficient spreading activation as well as sufficient speed of processing are required for the production and comprehension of referring expressions in discourse.

4 Results of the simulation

To evaluate the performance of the model, it was subjected to simplified but similar production and comprehension tasks as the participants in Wubs et al.'s (2009) study. We tested the model's production and comprehension of stories consisting of six sentences about two characters of the same gender, similar to the stories in Figure 5.

a. Story with topic shift	**b. Story without topic shift**
1. The pirate is playing with a ball.	1. The pirate is playing with a ball.
2. He kicks the ball in the water.	2. He kicks the ball in the water.
3. He tells a knight what happened.	3. He tells a knight what happened.
4. The knight rescues the pirate's ball.	4. The pirate asks the knight for help.
5. The knight gives the ball to the pirate.	5. The pirate gets a net from the knight.
6. He is happy.	6. He rescues the ball.

Figure 5: Examples of stories with and without a topic shift (adapted from Wubs et al., 2009). The references to the two characters are underlined.

The stories differed in whether they do or do not contain a topic shift. In stories with a topic shift (Figure 5a), the second character becomes the new topic halfway through the story. In stories without a topic shift (Figure 5b), the first character remains the topic throughout the story. The final sentence of each story contains a potentially ambiguous subject pronoun.

In the *production* task, the model was presented with five sentences as in Figure 5, to bring the computational model in a state it would have after producing these five sentences. The model was then asked to produce a referring expression for the firstly introduced character, which was not the current discourse topic. In the *comprehension* task, the model was presented with six sentences as in the stories of Figure 5. The referent that the model selected as the antecedent of the ambiguous pronoun was registered. To study development over time, we included training sessions for the model, so that it could acquire perspective taking. In these trials the model was not presented with stories, but with either a pronoun or full NP (in comprehension) or a topic or non-topic referent (in production).

Thus, in each run, the model alternated between experimental sessions (with one production task consisting of 4 trials and one comprehension task consisting of 8 trials) and training sessions (which consist of a total of 400 production and comprehension trials between two subsequent experimental sessions). Learning was turned off during the experimental sessions to allow for the repeated presentation of the same experimental stimuli (cf. McClelland, 1995; van Rij et al., 2010; van Rijn et al., 2003). Each run consisted of four experimental sessions, with interleaved training sessions. To simulate the difference in WM between children and adults, no activation spread from the subject to other linguistic structures in the first two experimental sessions (resulting in low WM simulations), whereas spreading activation was enabled in the last two experimental sessions (resulting in high WM simulations). We will first discuss the learning of perspective taking during the training and the effect of the WM manipulation in the production and comprehension task, and then evaluate the model's performance in the experimental sessions. The model's latter performance will be compared to the performance of the children and adults in Wubs et al.'s study.

4.1 Simulating the acquisition of perspective taking

During training, the model was presented with comprehension and production trials in random order. The input for comprehension trials was the type of discourse referent for which a expression had to be found, either 'topic' or 'non-topic'. The input for production trials was a type of referring expression for which an interpretation had to be found, either a full NP or a pronoun. The input for the trials was randomly determined. In production, the model had to select a referring expression for a 'topic' referent in 75% of the trials, and for a 'non-topic' referent in the remaining 25%, thus reflecting the intuition that people tend to refer to the topic more often than to a non-topic. In comprehension, the model was presented with pronouns and full NPs with equal probabilities and had to determine whether the expression referred to a 'topic' referent or a 'non-topic' referent. For practical reasons we simplified the training trials and did not present the model with entire discourses during training. We simply provided the model with information about the discourse status of the two referents (i.e.,

'topic' or 'non-topic'). Therefore, during training the acquisition of perspective taking was not influenced by the WM manipulation.

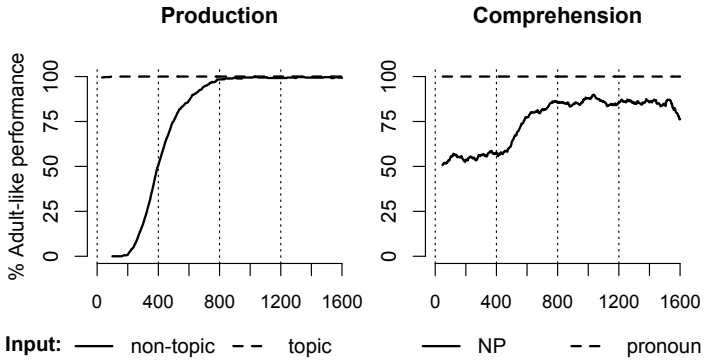


Figure 6: Learning to use and interpret referring expressions on the basis of the training data, mean of 15 simulations. Moving average (25 trials) of the percentage adult-like performance per number of training trials. The four vertical lines indicate the four experimental sessions alternating the training sessions of 400 trials each.

Figure 6 displays the percentages of adult-like performance on the training trials. In production, the dashed line, which is at ceiling from the start, indicates that the model correctly produces a pronoun for reference to a topic from the start, as pronouns are the preferred form according to the constraints of the grammar. As a consequence of this preference, however, the model initially also uses pronouns for reference to a non-topic, showing non-adult like performance. To reach adult-like performance with non-topics, the model needs more training trials. Recall that the selection of a full NP requires bidirectional optimization, which can only be performed after sufficient experience. In comprehension, a similar pattern can be observed. As the constraints of the grammar specify the interpretation of pronouns, pronoun comprehension is adult-like from the start. However, the model needs more trials to learn that a full NP signals a topic shift. Initially, the model randomly chooses between a topic and a non-topic interpretation because the grammar does not restrict the meaning of a full NP. Gradually, the model learns that a full NP is interpreted as reference to a non-topic referent because a pronoun would have been used if the speaker had intended reference to a topic referent.

Two mechanisms drive the learning of adult-like production and comprehension: the production compilation mechanism (introduced above, Taatgen & Anderson, 2002) and internal feedback. Recall that the production compilation mechanism shortens the processing time by combining production rules that are repeatedly performed in sequence into a single, more efficient production rule. During training, a success was registered if the candidate selected by the model was identical to the output resulting from bidirectional optimization. In other words, a success was registered if the model came up with the same solution an adult speaker would have selected according to the linguistic analysis. This success increases the chances

of using the same steps at the next run by updating the parameters associated with the production rules (Hendriks et al., 2007; Taatgen & Anderson, 2002; van Rij et al., 2010; van Rijn et al., 2003). As a result, sequences of production rules that lead to a success are more likely to undergo production compilation, thus speeding up the learning process.

Not every simulation results in 100% adult-like performance on the comprehension of full NPs. Because the constraints of the grammar do not restrict the interpretation of full NPs, the comprehension of full NPs generally takes more time than the production of full NPs for reference to non-topics, which also requires bidirectional optimization. To decide between a topic referent and a non-topic referent in comprehension, the model tries to retrieve another constraint. However, this takes extra time. Some simulations cannot complete the bidirectional optimization process that is necessary for the comprehension of full NPs within the available time and, as a result, no feedback is received. With unlimited time for interpretation, every simulation would reach 100% adult-like performance.

To summarize, the results of the training sessions in our simulations show that the model gradually acquires the efficiency to complete bidirectional optimization as a result of linguistic experience (cf. van Rij et al., 2010). Our simulations thus suggest that learning to take into account the opposite perspective in communication, a Theory of Mind-like process, is shaped by the frequency of the linguistic forms in the input, which determines how much time it takes to learn more efficient production rules. In the next section we discuss how WM capacity affects the discourse status of referents in our model and consequently may affect the input for the process of perspective taking.

4.2 Simulating working memory differences

The model not only received training materials, but was also presented with two experimental tasks (a production task and a comprehension task) at four points in the simulations (see Figure 6). In contrast to the training trials, the experimental trials were not simplified and the model was presented with entire discourses which either did or did not contain a topic shift. Recall that in the *production* task, the model produces a referring expression after processing the first five sentences of a story. In the *comprehension* task, the model identifies the antecedent of the potentially ambiguous pronoun in the sixth sentence of the story. For both tasks, it is required that the model on the basis of the first five sentences first determines what is the discourse topic. Therefore, we first discuss how the model's selection of the discourse topic is influenced by its working memory capacity. After that, we evaluate the model's results on the simulated production and comprehension tasks.

Figure 7 illustrates how spreading activation affects the choice for one of the referents as the current discourse topic. A low WM simulation and a high WM simulation are presented with the same story with a topic shift. The activation of the two discourse referents in the story is measured in both simulations during the processing of the five sentences. When the model is presented with the first referring expression in the story, a new discourse chunk is created that connects information about the current discourse with lexical information about the referent. Importantly, as can be seen in Figure 7, the activation of the discourse chunk,

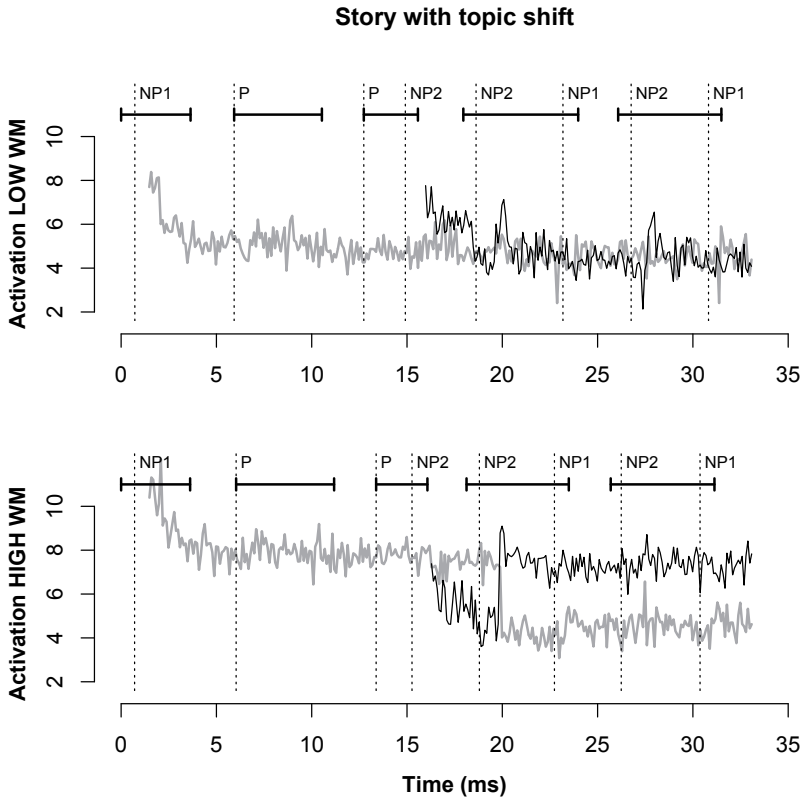


Figure 7: Activation of the two referents over time (s) in a story with topic shift in a low WM simulation (top) and a high WM simulation (bottom). The horizontal lines represent the duration of the first five sentences of the story, whereas the vertical lines depict the start of a referring expression (NP1 represents a full NP unambiguously referring to referent 1, NP2 unambiguously refers to referent 2, and P is a pronoun). The gray activation line represents the activation of the firstly introduced referent; the black activation line represents the activation of the second referent.

represented by the thick gray line, is higher in the high WM simulation than in the low WM simulation. The first referent is introduced in subject position. Therefore, in the high WM simulation the spreading activation from the subject boosts the activation of this discourse referent. In Sentence 2 and Sentence 3, a pronoun is used for the first referent. As there is only one referent available in the current discourse, the low WM and high WM simulation will always select this referent as the antecedent of the pronouns. The second referent is introduced with a full NP in the third sentence and this referent becomes the subject in the next two sentences. In the high WM simulation, the spreading activation from the subject causes

an increase in activation of the second referent, represented by the thin black line in Figure 7. In addition, it causes a decrease in activation of the first referent because this referent does not receive activation anymore from spreading activation. In the low WM simulation, the activation of both referents fluctuates around the same activation level because no referent receives activation from spreading activation. The different frequencies of mentioning of the first and second referent cause only a small difference in activation.

After processing all five sentences, the model retrieves the referent with the highest activation at that point in the discourse as the current discourse topic. Due to spreading activation, the referent that was the subject of the previous sentence will be selected as the discourse topic in the high WM simulation. In contrast, in the low WM simulation the two referents are equally likely to be selected as the discourse topic. For stories without a topic shift (not shown here), the activation pattern for the low WM simulation looks the same as for stories with a topic shift. In the high WM simulation, however, the first referent in stories without a topic shift will be selected more often as the discourse topic than in stories with a topic shift. This is because in stories without a topic shift, the activation of the first referent remains high throughout the story, as the first referent is the subject of each sentence.

Note that for simplicity, we modeled WM capacity as a discrete phenomenon. The low WM simulation does not spread activation from the subject at all, whereas the high WM simulation spreads a large amount of activation. Spreading activation from the subject reflects the capacity to keep information about the grammatical roles in the previous sentence accessible. Of course, we expect this capacity to be gradually acquired through maturation.

Our simulations show that spreading activation affects the activation of discourse referents and thus provides an explanation for the way differences in working memory may alter the use and interpretation of pronouns in discourse. In the previous section, we explained how the model's referential choice was altered by its ability to take into account the listener's perspective. The combination of these two factors, the amount of working memory capacity and perspective taking, may explain the differences between children's and adults' performance with referring expressions. In the next two sections, we evaluate the performance of our model on the production and comprehension tasks.

4.3 Simulating the production task

The simulated production task is a simplified version of the storybook task of Wubs et al. (2009) and consisted of four different experimental trials. As mentioned earlier, the model is presented with the first five sentences of a story to give it a background, and is then asked how it would continue the story. Thus, in every trial, the model is presented with a story with a topic shift presented word by word. After processing these five sentences, the model's goal is to produce a referring expression to re-introduce the first character, which is not the current discourse topic. The first step in selecting the referring expression is to determine whether the given referent is the current discourse topic (see Figure 3). This information is the input to the second step, an optimization processes that selects a referring expression and checks the recoverability of its meaning. The output of the model is either a pronoun or a full

NP. After 600 training trials and two experimental sessions, spreading activation is increased to simulate an increased WM capacity due to cognitive development. Figure 8 compares the model's performance with the findings of Wubs et al. The model captures the main effects of the data. That is, both children and the low-WM simulation overuse pronoun, and adults and the high-WM simulation prefer full NPs.

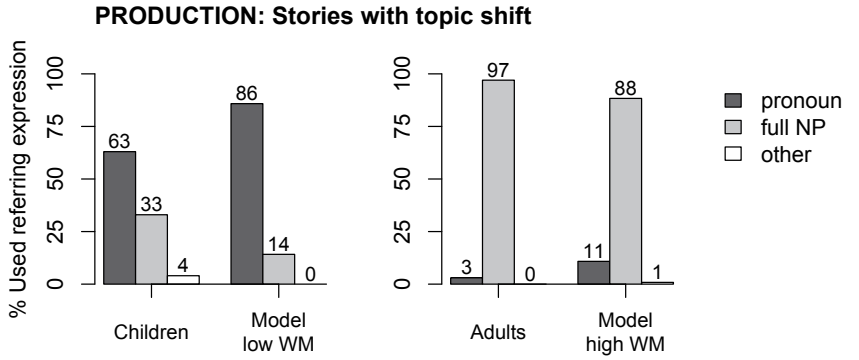


Figure 8: Production of referring subjects. The performance of the child and adult participants in the experiment of Wubs et al. (2009) is compared with the performance of the low-WM and high-WM simulations (mean of 15 simulations).

The overuse of pronouns in the low-WM simulations (86% of the trials; mean of 15 simulations) is caused by its low amount of spreading activation (57% of the trials) and its limited processing speed (29% of the trials). In 57% of the trials the intended referent is incorrectly considered to be the current discourse topic, as there is no spreading activation to increase the activation of the subject of the previous sentence. If the referent were the discourse topic, the adult choice would indeed have been a pronoun. So in these cases, the error resides in the first step of determining the discourse status of the referent. On the other hand, in 68% of the trials where the low WM simulations correctly consider the intended referent to be a non-topic (i.e., in 28% of all trials), the model chooses a pronoun for reference. In these trials, the error results from the later steps of the production process, as the model is not fast enough yet to take into account the listener's perspective. During the training sessions the model gradually acquires more efficient production rules, so that it can use this process of perspective taking in the production tasks to determine whether the selected referring expression is interpretable for a listener. In the high WM simulations, in 11% of the trials the process of perspective taking is not completed and consequently a pronoun is chosen for reference. Thus, even if the model has sufficient WM capacity, for adult-like production Theory of Mind-like perspective taking is required.

4.4 Simulating the comprehension task

In each experimental session of our simulation of the comprehension task, the model receives four stories with a topic shift and four stories without a topic shift. Each time the model encounters a pronoun, it starts the resolution process by retrieving the discourse topic and evaluating on the basis of the grammar whether the optimal interpretation of this pronoun is the discourse topic or a non-topic. We coded whether the pronoun in the final sentence of the story is interpreted by the model as referring to the first or the second referent. Figure 9 compares the model's performance with the findings of Wubs et al. (2009).

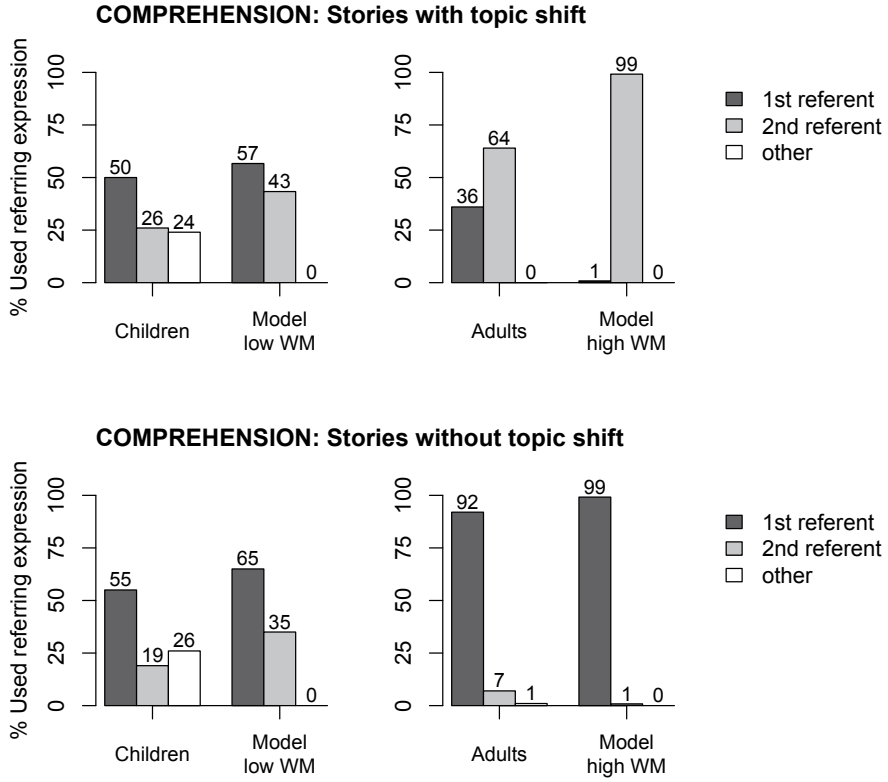


Figure 9: Comprehension of referring subjects. Performance of the child and adult participants in the experiment of Wubs et al. (2009) is compared with performance of our low-WM and high-WM simulations (mean of 15 simulations). The upper plots display the performance on stories with topic shift, the lower plots display performance on stories without a topic shift.

The model captures the two main effects in the children's data: (a) the low-WM simulation shows a small preference for the first referent, and (b) the low-WM simulation shows no

difference in performance between the two discourse conditions. The model further captures the pattern shown by adults: in the high WM simulations the model displays a large difference in performance between the two discourse conditions.

In contrast with the model's performance on the production task, its performance on the comprehension task is mainly determined by the WM manipulation. For the *comprehension* of a pronoun, perspective taking is not required as the constraints of the grammar (in particular the constraint PROTOP) already specify the discourse topic as the antecedent. Our simulations thus suggest that an adult-like *comprehension* of pronouns does not require Theory of Mind-like perspective taking but merely sufficient WM capacity.

As can be seen in Figure 9, the model captures the main effects found in the empirical data. However, whereas the high-WM simulation almost always selected the second character as the antecedent of the pronoun in the comprehension stories with a topic shift, the adults still chose the first character in a third of the cases. We can think of several explanations for this pattern. First, it is conceivable that the first mention of a referent boosts the activation of this referent in such a way that this referent remains more active than subsequently introduced referents. A second explanation, put forward by Koster et al. (2011), is that adults' frequent selection of the first referent in the topic shift condition may have been caused by the structure of the sentences in the construed stories. Although the stories with a topic shift keep the first referent active by continued reference also after the topic shift, to match the production stories, it may have been more natural to omit mention of the first referent when the second referent has become the new topic. Further research is needed to identify the exact cause of this additional effect in the adult data, which is also the reason why we decided against implementing one of these strategies in our model. Most importantly however, this additional effect does not affect our explanation of the experimental results.

Another difference between the model and the empirical data is the fact that the model does not account for the 'other' responses that children sometimes gave in the comprehension task. These 'other' responses included references to previous stories, and may have been triggered by difficulties in processing the current input but might also be genuine references to the most activated discourse referent. However, as our model processed each story independently from all other stories, it had not access to other characters as referent for the ambiguous pronoun.

5 General discussion

To investigate how linguistic constraints, discourse structure, and the speaker's estimation of the listener's interpretation determine the speaker's choice of referring expression, we developed a computational ACT-R model of the acquisition and use of referring expressions. On the basis of this model we propose that for adult-like performance on the production and interpretation of referring subjects not only sufficient WM capacity is necessary, but also sufficient speed of processing to perform the Theory of Mind-like process of perspective taking. Below, we discuss the relation between referential choice and linguistic experience, Theory of Mind and working memory capacity, respectively.

5.1 Relation between referential choice and linguistic experience

In our cognitive model, we implemented the speaker's task as a production step followed by a comprehension step. This implementation was based on a linguistic account of the use of referring subjects (Hendriks et al., 2008). This account holds that speakers not only select the form that is optimal according to the constraints of the grammar, but also consider the listener's perspective. This second step is needed to determine whether the intended meaning is recoverable from the selected form. The hypothetical listener uses the same grammar as the speaker, but selects the best meaning for the given form, rather than the best form for a given meaning.

Simulations of our computational model showed that this process of bidirectional optimization becomes more efficient when it is repeatedly used. Eventually, the process becomes so efficient that bidirectional optimization can be performed within the limited time available. Because this process depends on frequency of use, the frequency distribution of the linguistic input determines the speed of acquisition of bidirectional optimization. This explains why children, who initially produce unrecoverable pronouns, acquire the adult-like use of referring expressions with linguistic experience.

5.2 Relation between referential choice and Theory of Mind

In our model, production and comprehension are implemented as gradually specializing instantiations of bidirectional optimization. Given the similarity to Theory of Mind-like processes, it is expected that possessing these skills is required for the mature use of referring expressions (cf. Arnold et al., 2009; Gundel et al., 2007). When children possess Theory of Mind skills, they should in principle be able to apply this ability in taking into account the perspective of the conversational partner in referential choice. However, this complex process initially takes too much time to complete during on-line sentence production and comprehension. Gradually, however, the process of bidirectional optimization becomes more efficient through linguistic experience. Consequently, the model predicts that the application of Theory of Mind skills to language is not acquired through maturation, but rather on the basis of the input frequencies of the relevant linguistic forms.

5.3 Relation between referential choice and working memory capacity

In previous studies (Hendriks et al., 2008; Koster et al., 2011), correlations were found between WM capacity and elderly adults' and children's use of referring expressions. In our model, WM capacity was implemented as the amount of spreading activation from the sentential subject to all associated discourse referents.[3] As the subject of the preceding sentence is an important factor in determining the current topic, the subject information has to be retained in the goal buffer across sentences. In that respect, our implementation of spreading activation differs from the spreading activation account of Reitter (2008) that limits spreading activation to within-sentence effects.

Our model predicts that it is more difficult to determine the discourse topic for language users with a low WM capacity than for language users with a high WM capacity. If WM capacity is high, the discourse referent associated with the subject of the previous sentence receives an extra amount of activation due to spreading activation from the subject. As a result, a high-WM simulation is more likely to select the subject of the previous sentence as the discourse topic. On the other hand, if WM capacity is low, the discourse topic is determined on the basis of frequency and recency only. Because the two discourse conditions in the experimental task did not differ in the frequency of the referents, the low-WM simulation displayed similar performance on these stories. So WM capacity determines whether the model can take into account the discourse structure. This prediction is in line with the working memory-account of Daneman and Carpenter (1980). They argue that readers with a smaller reading span, a measure correlated with working memory capacity, are less likely to have prior information active in working memory, and hence these readers will be less successful or take more time to compute the referent of a pronoun. Results of a study with adult participants (van Rij, van Rijn, & Hendriks, 2011) provide additional support for our model's prediction: In a dual-task experiment in which one of the tasks was a comprehension task similar to the one in the present study, participants were less likely to consider the subject of the previous sentence as the discourse topic when the secondary task induced a high WM load, in comparison with a secondary task that was less demanding. That is, when adults' working memory was taxed, their comprehension of pronouns in discourse became more child-like.

5.4 Integration of discourse-based and listener-based approach

Even if our cognitive model has a high WM capacity, it will still need perspective taking to block the production of unrecoverable pronouns for non-topic inputs. On the other hand, if the model is able to apply perspective taking successfully but has a low WM capacity, it will have difficulty determining the discourse topic. If an incorrect choice is made for the topic and the referent is incorrectly taken to be the discourse topic, the model will produce an unrecoverable pronoun too. This may explain why elderly adults with a low WM capacity, who have had sufficient linguistic input, nevertheless produce unrecoverable pronouns (cf. Hendriks et al., 2008). For the correct use of referring expressions, therefore, both sufficient WM capacity and sufficient processing speed are necessary. Our model thus integrates the discourse-based approach to referential choice (Ariel, 1990; Givón, 1983) with the listener-based approach (Gundel et al., 1993), while providing a truly cognitive basis for the term 'cognitive status' employed in this linguistic work. Furthermore, the results of our model contribute to the ongoing debate as to whether speakers and listeners coordinate their actions (e.g., Brennan & Hanna, 2009; Clark, 1996; Keysar et al., 1998; Shintel & Keysar, 2009; Wardlow Lane & Ferreira, 2008): In our model, speakers aim at taking into account the hypothetical listener, but are not always able to do so due to processing limitations.

5.5 Further predictions of the model

Besides fitting existing experimental data, the advantage of a cognitive model is that it can generate precise predictions regarding language users' performance in new linguistic tasks. Wubs et al. tested the production of stories with topic shift and the comprehension of stories with and without topic shift. In addition to these three situations, which we could simulate using the experimental results from the Wubs et al. study, we looked at the model's performance on the production of similar stories *without* a topic shift, which were not investigated by Wubs et al. In such stories, the first referent is the topic throughout the story. After hearing a discourse consisting of five sentences, the model's task is to produce the optimal expression for referring to the first character. On the basis of the grammar, adults are expected to use a pronoun. However, the low WM simulation predicts a temporary decrease in the use of pronouns by children, as a result of the interaction between learning to apply perspective taking and a low WM capacity (see Figure 10). When children incorrectly consider the second character to be the topic, and at the same time happen to succeed in perspective taking, they will be overly specific and produce a full NP. This prediction has not been tested yet.

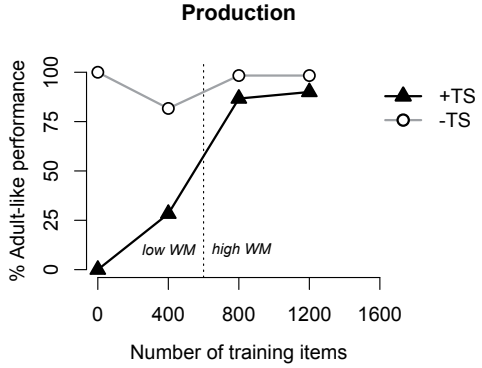


Figure 10: Predictions of our ACT-R model for the production of stories without topic shift, compared with the performance on the production task testing stories with topic shift (mean of 20 simulations). The symbols show the percentages adult-like performance in the four experiments (mean of 20 simulations): ▲ represent stories with a topic shift (+TS), ❍ represent stories without a topic shift (-TS). After 600 training trials, the model's WM capacity is increased (indicated by the dashed line). Note that the performance on these experimental sessions displays a similar pattern as the acquisition of perspective taking shown in Figure 6. As Figure 6 depicts the acquisition of perspective taking in training sessions, this indicates that performance depends on the model's use of perspective taking.

Another prediction of the model concerns the comprehension of much simpler stories with mainly intransitive sentences that have only one referent per sentence (e.g., "The knight wants to play with the ball. He..."), as compared to the transitive sentences we modeled previously (e.g., "The knight gives the ball to the pirate. He ..."). In production, the model's task is to determine the optimal referring expression for re-introducing a previously men-

tioned referent. Because the low-WM simulation has a slight preference for the most recently mentioned second referent as the topic, the model produces more full NPs than it did in the production stories discussed earlier. In comprehension, the model's task is to determine the optimal antecedent of the pronoun in the final sentence of the story. In stories *with* a topic shift, the low-WM simulation shows a slight preference for the most recently mentioned second referent. In stories *without* a topic shift, the low-WM simulation will correctly select the first character as the referent of the pronoun because the first character is much more frequent than the second character. Thus, the computational simulations predict that children will perform better on pronoun resolution in simple discourses consisting of intransitive sentences. Although this prediction might seem straightforward, a more stringent prediction is that performance in these conditions will not be correlated with WM capacity.

6 Conclusion

In this paper, we presented a computational model of children's and adults' production and comprehension of referring subjects. By modeling the speaker's task as a step of production followed by a step of comprehension, the model is able to explain children's non-adult-like production of unrecoverable pronouns. Due to insufficient processing speed, children are not yet able to perform the consecutive steps of this Theory of Mind-like process of perspective taking in an adult-like way. Based on the simulations of our model, we argue that the speaker's consideration of the listener's perspective is crucially dependent on speed of linguistic processing, whereas the speaker's identification of the discourse topic is dependent on sufficient WM capacity. These two cognitive measures may develop independently: speed of linguistic processing was shown to increase as an immediate effect of linguistic experience, whereas WM may develop with age. The integration of a theory of grammar and discourse in a cognitive model allows us to provide cognitively plausible explanations for previous experimental results. In addition, this approach is able to generate precise predictions regarding language users' performance in new linguistic tasks. This may provide a fruitful way to gain new insights into the complex interaction between linguistic constraints and various cognitive factors in language processes.

Notes

[1] For modeling the use of referring objects (van Rij et al., 2010), two steps were sufficient. This corresponds to strong bidirectional optimization in an OT framework. For modeling the use of referring subjects, on the other hand, sometimes more than two steps are necessary. This corresponds to the recursive version of bidirectional optimization, weak bidirectional optimization (cf. Blutner, 2000).

[2] Initially, a single process of unidirectional optimization takes at least 250 milliseconds because at least four production rules (which each take 50 milliseconds to fire) are necessary to evaluate two candidates on the basis of one constraint. This is a relatively

long time in on-line sentence production and comprehension, considering the fact that speech is typically produced at a rate of 3 or 4 syllables per second.

[3] In our model, we implemented processing efficiency and WM capacity as two independent cognitive constraints. However, the effects of the two mechanisms interact. Given that WM capacity is considered to be the amount of source activation (cf. Daily et al., 2001), a higher WM leads to a higher activation of chunks. If chunks have a higher activation, the retrieval time is shorter (Anderson et al., 2004). As a result, an increase in WM capacity leads to an increase in processing speed, which enhances the effect of processing efficiency. Thus, although linguistic experience and WM capacity are two independent factors, both influence processing speed and consequently affect the onset of adult-like performance.

chapter 4

How WM load influences linguistic processing in adults:

A computational model of pronoun interpretation in discourse

This manuscript was submitted as:
Van Rij, J., Van Rijn, H., & Hendriks, P. (submitted). How WM load influences linguistic processing in adults: A computational model of pronoun interpretation in discourse. Manuscript, September 2012.

Abstract

This paper presents a study of the effect of working memory load on the interpretation of pronouns in different discourse contexts: stories with and without a topic shift. We discuss a computational model (in ACT-R, Anderson, 2007) to explain how referring expressions are acquired and used. On the basis of simulations of this model, new predictions were generated with respect to the influence of WM load in adults, suggesting that WM constraints only affect pronoun resolution in stories with a topic shift, but not in stories without a topic shift. This latter prediction was tested in an experiment. The results of this experiment support the prediction that WM load has specific effects on adults' off-line interpretation of referring expressions in discourse.

1 Interpretation of subject pronouns

Listeners interpret referring expressions such as pronouns by using information from the preceding linguistic discourse. However, representing and accessing this discourse information may come with certain costs. In this study we investigate the influence of working memory load on adults' interpretation of pronouns in discourse.

Table 1 shows two short stories about two characters of the same gender. In the stories, the final sentence starts with a potentially ambiguous pronoun (*he*), which can refer to both male characters in the given linguistic context (*Eric* and *Philip*).

A. Story with topic shift	**B. Story without topic shift**
1. Eric / gaat / voetballen / in de sporthal. 'Eric is going to play soccer in the sports hall.'	1. Eric / gaat / voetballen / in de sporthal. 'Eric is going to play soccer in the sports hall.'
2. Philip / vraagt / Eric / om mee te rijden/naar de training. 'Philip asks Eric to carpool to the training.'	2. Eric / vraagt / Philip / om mee te rijden/naar de training. 'Eric asks Philip to carpool to the training.'
3. Philip / haalt / Eric / na het eten / met de auto op. 'Philip picks up Eric after dinner by car.'	3. Eric / haalt / Philip / na het eten/met de auto op. 'Eric picks up Philip after dinner by car.'
4. Hij / voetbalt / al twintig jaar. 'He has played soccer for twenty years.'	4. Hij / voetbalt / al twintig jaar. 'He has played soccer for twenty years.'
Question: Wie voetbalt al twintig jaar? / 'Who has played soccer for twenty years?'	

Table 1: Example of a story in Dutch with (A) and without (B) topic shift.

Generally, pronouns are used to refer to the most salient character or entity (i.e., the *topic*) in the linguistic discourse (a.o., Ariel, 1990; Gundel et al., 1993). In contrast, full noun phrases or proper names are used to refer to less salient characters or to introduce new characters. Different factors have been found to affect the interpretation of pronouns (see Arnold, 1998, for a review), among others the grammatical role of potential referents. As the subject of the previous sentence is likely to be the current topic (Grosz et al., 1995), listeners will often interpret a pronoun as referring to the previous subject (a.o., McDonald & MacWhinney, 1995; R. Stevenson, Crawley, & Kleinman, 1994).

Koster et al. (2011) presented children (age 4-6) and adults with prerecorded stories, of which the last sentence started with a potentially ambiguous pronoun. In half of the stories the topic shifted halfway by changing the grammatical role of the characters (similar to the stories in Table 1). In agreement with the literature, they found that adult listeners prefer to interpret the ambiguous subject pronoun as referring to the subject of the previous sentence (*Philip* in Table 1a, and *Eric* in Table 1b). Children, however, showed a general preference for the first character of the story as the referent of the pronoun (*Eric* in Table 1a and 1b), and did not distinguish between stories with and without topic shift. Interestingly, Koster et al. report that children with a higher auditory working memory capacity performed more adult-like in

their pronoun use.

We implemented a computational model within the cognitive architecture ACT-R (Anderson, 2007) that simulates the acquisition of referring expressions in discourse (van Rij, van Rijn, & Hendriks, submitted*a*). The computational simulations of our ACT-R model suggest that limited WM capacity may cause decreased performance on children's choice and interpretation of referring expressions in discourse, thus providing a detailed, mechanistic explanation of the empirical correlation reported by Koster et al. (2011). Our ACT-R account is in line with previously proposed computational models in different frameworks that explain the relation between WM capacity and language processing, such as CC READER (Just & Carpenter, 1992) or 4CAPS (Just & Varma, 2007). In these models, WM capacity is implemented as a limited amount of activation, which is different for different individuals. If more capacity is required for processing or storage than is available, this will result in longer processing times or retrieval errors (e.g., Daneman & Carpenter, 1980). However, alternative accounts have been put forward to explain the relation between WM capacity and language processing. For example, MacDonald and Christiansen (2002) propose that differences in WM capacity are differences in skill that arise from variations in exposure to the language and biological differences. Alternatively, Ericsson and Kintsch (1995) and Kintsch (Construction-Integration model; 1998) argue that WM capacity reflects the efficiency of retrieval cues present in the active portion of working memory. In their view, more efficient cues result in more accurate and faster retrieval of information from declarative memory.

In this paper we show that our ACT-R model makes specific predictions about adults' linguistic performance under different conditions of WM load. Based on simulations of our ACT-R model, we expect adults' interpretation of subject pronouns in discourse to be influenced by the amount of WM capacity available for interpretation. We present empirical evidence supporting this novel prediction.

2 Cognitive model

In a previous study, we implemented a computational model within the cognitive architecture ACT-R (Anderson, 2007) that simulates children's acquisition of referring expressions in discourse (van Rij et al., submitted*a*). In our implementation, we focused on the differences between potentially ambiguous pronouns (*he*) and more explicit referring expressions such as full noun phrases (*the soccer player*) or proper names (*Eric*). The cognitive model was shown to simulate the acquisition data reported in Koster et al. (2011). This paper presents new predictions, following from simulations of the cognitive model, for *adults'* interpretation of referring expressions in discourse. Before describing the simulations, we will give a short overview of the model.

2.1 Overview of the model

The model consists of two parts, namely the memory principles that determine the representation of the linguistic discourse, and the linguistic principles that underlie the choice

and interpretation of referring expressions. The model uses general memory principles to build a representation of the discourse during sentence comprehension. The saliency of discourse referents is represented by the activation of elements in memory. The discourse referent with the highest activation is taken to be the current discourse topic. In addition to this memory-based process that automatically constructs a discourse representation, the model uses a rule-based process to perform its task of producing or interpreting a referring expression. To *produce* a referring expression, the model evaluates possible forms, such as a pronoun or a full noun phrase, on the basis of linguistic principles. It then takes the perspective of the listener to select the expression that is not only economical to use, but also understandable for the listener in the current discourse. To *interpret* a referring expression, the model evaluates possible discourse referents, such as the current topic, based on linguistic principles. Then it takes the perspective of a speaker to determine whether a speaker would indeed have used the expression for the selected interpretation. The model thus integrates linguistic theories based on the accessibility of referents (e.g., Ariel, 1990) and those that assume a pragmatic reasoning process (e.g., Gundel et al., 1993).

Here, we focus on the interpretation of subject pronouns, which is mainly driven by the linguistic discourse and not by sentence-internal factors, in contrast to object pronouns (van Rij et al., submitted*a*). The linguistic discourse must be represented in the listener's memory. By modeling the memory principles determining the representation of the linguistic discourse, we can generate model-based predictions about children's and adults' interpretation of referring expressions in discourse.[1]

2.2 Simulations: The effect of WM capacity on pronoun interpretation

Simulating the task of Koster et al. (2011), the model is presented with stories with and without topic shift. The six-sentence stories are presented to the model word by word. During on-line processing, the model builds a (simplified) representation of the preceding discourse: every time a character is encountered in the story, that character is represented in the declarative memory. Each memory representation (referred to as *chunk*) has a certain amount of activation that reflects the saliency of that character in the current discourse (consistent with Foraker & McElree, 2007; but in contrast with Grosz et al., 1995; Gundel et al., 1993). The final sentence of each story starts with a potentially ambiguous pronoun. As the current discourse topic, and hence the referent of the pronoun, the model selects the chunk with the highest level of activation.

In ACT-R, the activation of a chunk reflects the chunk's history and the chunk's usefulness in the current context (see Appendix 1). The activation of chunks decays with time, but is increased when the chunk is retrieved. This base-level activation is dependent on the frequency of use (the more frequently used, the higher the activation) and the recency of the last retrieval from memory (the more recent the last retrieval, the higher the activation). In addition, spreading activation from other chunks can temporarily boost a chunk's activation: chunks that are currently being processed spread activation to other, connected chunks in declarative memory. As the amount of spreading activation determines the ability to main-

tain goal-relevant information, differences in spreading activation account for individual differences in working memory capacity (Daily et al., 2001).

In our model, the subject of the previous sentence spreads activation to all discourse elements associated with it (similar to the implementation of short-term syntactic priming of Reitter, Keller, & Moore, 2011). We manipulated the amount of spreading activation to explain the difference between children's and adult's interpretation of subject pronouns reported by Koster et al.. If the amount of spreading activation is high, the chunk representing the subject spreads a large amount of activation and discourse elements that are associated with the subject become more activated in comparison with other discourse elements. As a result, the model will retrieve the referent that was the subject of the previous sentence as the current discourse topic. On the other hand, if the subject spreads a small amount of activation, reflecting a low WM capacity, then there will be no effect on the discourse elements associated with the subject. In that case, the effects of frequency and recency will be the main determinants of referent retrieval.

Figure 1 shows the effect of WM capacity (i.e., the amount of spreading activation) on the activation of the two referents in the stories with topic shift presented to the model.

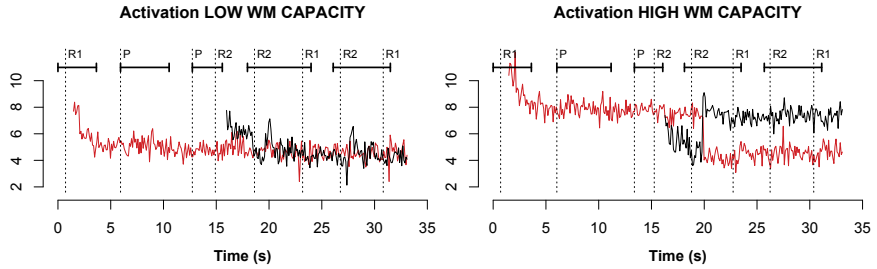


Figure 1: Activation of the two referents over time (s) in a story with topic shift in a low WM capacity and a high WM capacity simulation. The horizontal lines represent the duration of the first five sentences of the story, whereas the vertical lines depict the start of a referring expression (R1 and R2 represent expressions unambiguously referring to referent 1 and 2, and P is a pronoun). The red activation line represents the activation of the firstly introduced referent; the black activation line represents the activation of the second referent (adapted from van Rij, van Rijn, & Hendriks, submitted).

The second character is introduced in Sentence 3. The topic shift condition starts to differ from the no topic shift condition in Sentence 4, where the second character becomes the subject in the topic shift stories, but not in the stories without topic shift (cf. Koster et al., 2011). With a high WM capacity, the model selects the subject of the previous sentence as the referent of the pronoun in Sentence 6, because this discourse element clearly has the highest activation (Fig. 1, right). However, with a low WM capacity, the model will show a much-reduced preference for the second character as the referent of the pronoun, and often chooses the first character. Similarly to children's performance, the model's interpretation

of pronouns is not affected by grammatical role (Fig. 1, left).

2.3 New prediction: Effect of WM load on adults' interpretation of pronouns

On the basis of our model's simulations we propose that an individual's WM capacity influences the extent to which the grammatical structure of the previous sentence plays a role in resolving a potentially ambiguous subject pronoun. If this hypothesis is correct, we expect that adults show difficulties in detecting a topic shift when their WM is taxed by having to perform a memory task in parallel. This prediction follows directly from the ACT-R model: goal-relevant information spreads a proportion of the total spreading activation to other chunks in the declarative memory. If the number of sources from which activation is spread increases, the amount of spreading activation received by individual chunks decreases. In a high WM load situation, more information needs to be maintained in an activated state and as a result, the subject of the previous sentence spreads less activation to the discourse elements associated with the subject. Therefore, the model predicts that adult listeners or readers show more child-like performance in high WM load conditions: they will more often select the first character as the current discourse topic. In addition, as the level of activation determines the retrieval time, the model predicts that it will take more time to retrieve a discourse referent in a high WM load situation: due to spreading activation, the activation of discourse referents in a high WM load condition is lower.

3 Dual-task experiment

We performed a dual-task experiment to test our prediction that adult listeners will show difficulties in their comprehension of a topic shift if they have less WM capacity available (see for a similar dual-task manipulation in the domain of the processing of referring expression, Campana, Tanenhaus, Allen, & Remington, 2011). To manipulate WM load, participants were asked to perform a combined task: memorizing a sequence of digits for later recall while performing a moving-window task (Just, Carpenter, & Woolley, 1982).

3.1 Methods

3.1.1 Digit task

At the start of each trial, participants had to memorize a sequence of either three or six digits (low and high WM load conditions). Digits were shown for 1 second each in the center of a computer screen. The digits were pseudo-randomly chosen from 1 to 9, while ensuring that not all the digits were the same. After completing the moving-window task, the participants had to recall the memorized digits.

3.1.2 *Moving-window task*

After the presentation of the digits, participants had to read stories consisting of four sentences each (see Table 1), followed by a comprehension question. The sentences were presented one by one and were subdivided into smaller word clusters (indicated by slashes in Table 1). Using a typical moving-window paradigm (Just et al., 1982), only the letters of one single word cluster were visible at a time, and the other letters were replaced by a dot. By pressing a button, the participant could move the window to the next word cluster. After reading the four-sentence story, a question was presented with two answer alternatives. Participants had to press the corresponding button to answer the question. After answering the question, they had to type in the digits that were presented at the beginning of the trial. At the end of each trial, participants only received feedback on the digit task to ensure sufficient focus on the WM task. We collected different measures per trial: reading times per region, accuracy and reaction times for the questions and accuracy in reproducing the digits.

3.1.3 *Design*

Stories. Every story featured two characters of the same gender. The final sentence of each story started with an ambiguous subject pronoun *hij* ('he') or *zij* ('she'): the pronoun could refer to both characters, so that the only clue to the interpretation of the pronouns was the structure of the story. We designed two variants of every test story (see Table 1), in which we manipulated whether there was a topic shift or not. The topic shift is realized by placing the second character (*Philip*) in subject position in Sentence 2. If there was no topic shift, we expected participants to prefer the firstly introduced character as the referent of the ambiguous pronoun. If there was a topic shift, we expected participants to prefer the second character. At the end of every test story a question was presented to elicit the preferred interpretation of the ambiguous pronoun.

Lists. The presented materials were part of a larger experiment, in which we additionally tested two other variants of the stories. In total, 64 test stories were designed in four different variants. Four lists of 64 test stories (16 test stories per condition) were constructed to separate the different variants of the test stories. The lists also contained 128 filler items with the same structure as the test stories (32 filler stories per condition, the same for all lists). The filler stories were followed by a question about the first or second sentence of the story, to avoid reading strategies and to mask the goal of the experiment. Half of the filler questions asked about a character, the other half were what- or where-questions. Note that in contrast to the test questions, filler questions were not ambiguous and could be scored as right or wrong.

Here, we report on 2 x 32 test items, and the 64 filler items with the same two discourse structures. One test story (both variants) was removed from the data, because of a technical problem during presentation.

3.1.4 Procedure

Participants were randomly assigned to one of the four lists. The experiment consisted of two blocks: a low WM load block (3 digits) and a high WM load block (6 digits). The order of blocks was counterbalanced; within blocks the items were randomly distributed. Participants received instructions for each block. The experiment started with a practice trial suited for the current WM load condition.

3.1.5 Participants

Sixty-two first-year psychology students (18 men, 44 women; mean age 19) participated in the experiment in exchange for course credits. Five participants could not complete the experiment because of technical problems. Another 5 participants were excluded from data analysis, because they answered less than 75% of the filler questions correctly in the low WM load condition, and/or performed at chance level in one of the two types of filler questions. Data of 52 participants (15 men, 37 women) was used for the statistical analyses.

3.2 Results

In this section, we first discuss the performance on the digit task. We then present the off-line results on the linguistic task (i.e., the answers on the questions and the reaction times), followed by the on-line results on this task (i.e., the reading times).

3.2.1 Results on digit task

Participants made more errors on the digit task in the high WM load condition than in the low WM load condition (percentage correct trials: 3-digits=77.2%, 6-digits=52.0%), indicating that the 6-digit condition was indeed more difficult.

3.2.2 Off-line results on linguistic task

Answers. Figure 2 (left) shows the percentages of times that the subject of the previous sentence was chosen as the referent of the ambiguous pronoun. In addition, the percentages of correct answers on the filler questions (right) are given for comparison. Figure 2 shows that participants were sensitive to the topic-shift manipulation. In both WM load conditions, the subject of the previous sentence was preferred.

We examined the effects of *Topic shift*, *WM load*, *Trial position* (the trial position in the experiment) and *Session* (the first or the second block of the experiment) on the choice for the previous subject (yes or no) in the test items using logistic mixed-effects models (cf. Baayen, 2008). More complex models that included additional predictors such as list, the position of the subject answer on the screen (left or right), or the accuracy of repeating the memorized digits, did not show qualitatively different effects. In all the presented models, participant and item (all variants of a story were labeled as the same item) were included as crossed-random effects, with the maximum random effect structure that was justified by the data.

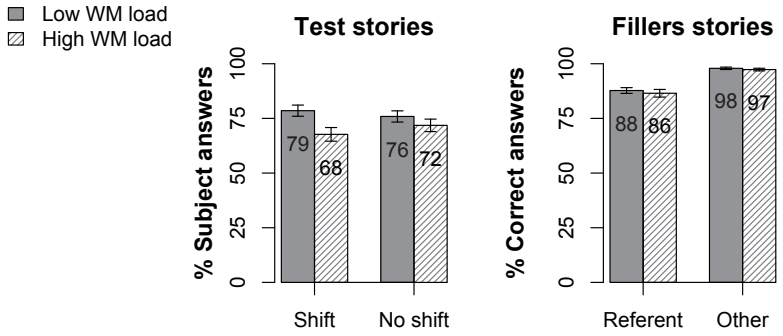


Figure 2: Left: Preference for the subject of the previous sentence as the referent of the pronoun for stories with and without topic shift (± 1SE); this would be *Philip* in the story with topic shift in Table 1, and *Eric* in the story without topic shift. Right: Percentage correct answers for filler questions asking about a referent or a non-referent (± 1SE).

We compared different models using a stepwise variable deletion procedure, starting with the complete interaction model (Baayen & Milin, 2010).

To explain the choice for the subject of the previous sentence (yes or no), the four-way interaction between *Topic shift*, *WM load*, *Trial position* and *Session* needed to be included in the model ($\chi^2(1)= 7.950$, p=.005). The full-interaction model (Appendix 2, Table A1) showed that in Session 1 the subject was selected less often in the high WM load condition than in the low WM load condition for stories with a topic shift (β=-0.676, z=-2.71; p=.007), in line with the assumption that decreasing WM capacity reduces the performance on pronoun resolution. However, *Trial position* attenuates this effect: near the end of the session, the difference between the WM load conditions became smaller (β=-0.336, z=-3.15; p<.001). A negative main effect of *Trial position* (β=-0.009, z=-2.98; p=.003) indicates that participants were less likely to choose the subject as referent near the end of the experiment. Note that these interactions between *WM load*, *Topic shift*, *Trial position* and *Session* are mainly driven by the topic shift condition: a mixed-effects model to explain the stories with a topic shift shows a significant three-way interaction between these predictors ($\chi^2(1)= 5.707$, p=.017). Crucially, no effects of *WM load*, *Trial position* or *Session* were found using the same analysis for the stories without a topic shift.

Reaction times. In the same way as we analyzed the choice of referent, we analyzed the log-transformed reaction times after excluding the short outliers (<=50ms; less than 1% of the data). We found a significant interaction between *Trial position* and *Session* ($\chi^2(1)= 7.783$, p=.005), indicating that participants became faster in answering during the experiment, but no significant effects of *WM load* and *Topic shift*. The best-fitting model additionally included a the predictor *Subject* ($\chi^2(1)= 11.597$, p<.001): Participants were faster when selecting the subject than when selecting the non-subject referent.

To summarize, we found that WM load affects the comprehension of stories with a topic

shift, but not the stories without a topic shift: participants less often select the subject of the previous sentence as the referent of the ambiguous pronoun in the high WM load condition. However, we did not find a difference in reaction times between the two types of stories, suggesting that the questions after stories with a topic shift are not more difficult to answer. These findings support our prediction that adults will show difficulties in processing a topic shift under higher WM load.

3.2.3 *On-line results on linguistic task*

Reading times. Before analyzing the reading time data we removed missing data (2%), short outliers (smaller than 50 ms, 19%) and used a log-transform to reduce the effect of the long outliers (cf. Baayen & Milin, 2010). The relatively large amount of short outliers was caused by a technical problem. As the outliers were equally distributed over the story conditions and the WM load conditions ($\chi^2(1)=0.925$, p>.1), it is unlikely that this influences our results in qualitative ways.

We analyzed the log-transformed reading times of Sentence 2 and Sentence 4, to determine whether WM load influenced the processing of the topic shift and the processing of the ambiguous pronoun.

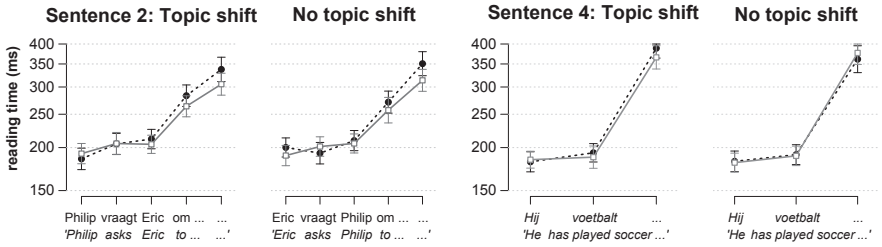


Figure 3: Reading times (± 1SE) of Sentence 2 (left) and Sentence 4 (right) in the low WM load condition (dashed line) and the high WM load condition (solid line). Note the logarithmic scale on the y-axis.

Sentence 2. Figure 3 (left) shows the reading times of Sentence 2 on a logarithmic scale. We expected an interaction in reading times between the topic shift conditions and the WM load conditions at the start of the sentence: The introduction of a new referent is expected to cause an increase in reading times, in comparison with repeated reference to the previous subject. However, this difference between the topic shift conditions is not expected for the high WM load condition, because the model predicts that a high WM load decreases the activation of the subject referent. To test these hypotheses, we analyzed the effects of *WM load* and *Topic shift* in the first and second region, which may reflect residual effects of retrieving the discourse referent after processing the word.

In the analysis of the first region, we included *Trial position*, *Session*, *Letters* (the number of letters of the word) in the linear mixed-effects model. The best-fitting model (Appendix 2, Table 2) contained significant interaction between *Trial position* and *Session* ($\chi^2(1)= 6.371$,

p=0.012), and between *WM load* and *Trial position* (χ^2(1)=18.829, p<.001). The latter interaction indicates that reading times became faster towards the end of the experiment in the high WM load condition (β=0.103, t=-2.42), although reading times were generally slower in the high WM load condition (β=0.001, t=2.09). There was no significant effect of *Topic shift* (β=-0.011, t<1).

In the analysis of the second region we included *Trial position*, *Session*, and *LogRT1* (the log-transformed reading times on the first region), to control for autocorrelation effects. However, we did not find any significant interaction. The main-effects model (Appendix 2, Table 3) did show a main effect of *Topic shift* (β=0.028, t=2.15, p=.032), indicating that readers needed more time to read the second region in the stories with topic shift, but no significant effect of *WM load* (β=0.015, t=1.13; p>.1).

Additional analyses that evaluate all reading times of the sentence to investigate the effects of *WM load* and *Topic shift* over time, showed significant effects of these predictors at the end of the sentence, but no interaction. These effects may point to a reasoning process or a task-specific strategy, which is beyond the scope of this study.

Sentence 4. Figure 3 (right) shows the reading times for the final sentence of the story, which started with an ambiguous pronoun. We analyzed this sentence using the analyses described before, to determine whether WM load influenced the retrieval of the referent of the pronoun. However, we did not find significant effects of *Topic shift* or *WM load*, nor an interaction between these two factors in the first two regions of Sentence 4.

To summarize, in Sentence 2 we found an effect of *WM load* on the first word and an effect of *Topic shift* on the second word, but no interaction between *WM load* and *Topic shift*. The prediction that more salient discourse referents give rise to shorter retrieval times was not reflected in the reading time data.

4 Discussion

We predicted, on the basis of our cognitive model, that adults would show more difficulties in processing a topic shift in higher WM load conditions. We performed a dual-task experiment to investigate this prediction. We hypothesized that as WM load increased, adult readers would show a significant decrease in their preference for the second character as the referent of a pronoun in the stories with a topic shift. In addition, we expected an increase in reading times in stories with a topic shift as a result of the topic shift, but we expected that this increase would diminish in the high WM load condition.

The off-line data support the model's prediction: participants selected the first character as the referent of the ambiguous pronoun significantly more often in the high WM load condition. No differences in reaction times were found, suggesting that the comprehension questions were equally difficult to answer for the two types of stories.

With respect to the reading times, we found an increase in reading times immediately after presenting a new referent in subject position. This indicates that readers expect to see the subject of the previous sentence again, instead of a new referent. However, we did not measure a significant interaction between WM load and story type. Different explanations

are possible for why this interaction did not reach significance, contrary to our expectations. It could be that WM load does not affect the processing of the sentence, but only affects the updating of the discourse representation with new sentence information. In that case, sentence wrap-up effects could have masked an interaction between topic shift and WM load. An alternative explanation is that the moving-window task is not suited for detecting the effect of WM load. It is reasonable to assume that the effect of WM load on topic shift is spread out over different regions and is thus more difficult to detect. ERP studies provide support for this explanation, because for unexpected noun phrases readers show an ERP effect 300-600 ms after the determiner of the unexpected noun phrase (Otten & van Berkum, 2009), which is much longer than it took participants in our experiment to read one region (approximately 200 ms).

The link between WM capacity and language processing is not new. Within the context of ACT-R, R. L. Lewis and Vasishth (2005) have explained difficulties in sentence processing, which have been attributed to WM load, by ACT-R's fluctuating activation and similarity-based interference in the retrieval of chunks. Our account is also in line with previously proposed accounts that consider WM capacity as a limited source of activation that is different for different individuals (a.o., Daneman & Carpenter, 1980; Just & Carpenter, 1992; Just & Varma, 2007). Alternatively, Ericsson and Kintsch (1995) argue that WM capacity reflects the efficiency of constructed representations in memory (Kintsch, 1998, Chapter 7), which is only consistent with our data under the assumption that higher WM load changes the representations that are constructed in memory. In other words, higher WM load may reduce the probability that the grammatical role is encoded in the discourse representation of a referent. This is subtly different from our implementation, since instead of changing the discourse representation itself, our account suggests that WM load affects the contribution of the previous subject to the activation of referents but does not change the internal representation. Furthermore, our data shows that WM load only affects the interpretation of stories with a topic shift, which is difficult to explain in terms of individual variation in exposure to the language and biological differences (as MacDonald and Christiansen, 2002, propose).

To conclude, we found that WM load can affect adults' interpretation of ambiguous pronouns by reducing the influence of the discourse context. This specific effect of WM load on pronoun interpretation was predicted by our cognitive model. Our model implies that readers or listeners without sufficient WM capacity rely more on the base level activation of discourse elements, instead of using information about the grammatical roles of the referents in the previous sentence. This study thus shows that language comprehension can be different for adults in different situations, and that adults may become more child-like under high WM load.

Notes

[1] The code of the discussed ACT-R model can be retrieved from: http://www.let.rug.nl/jacolienvanrij/wm-model.html

Appendices

1 Activation formula of ACT-R

In ACT-R (Anderson, 2007) the activation of chunk *i* is defined by: $A_i = \ln(\sum_{k=1}^{n} t_k^{-0.5}) + \sum_{j=1}^{m} W_j S_{ji} + \epsilon_i$, with *n* being the number of presentations of chunk *i*, and t_k the time since the *k*th presentation, *m* the number of chunks that are connected with chunk *i*, W_j the amount of activation that is spread from chunk *j*, S_{ji} the strength of association between *j* and *i*, and ϵ_i noise. The activation of a chunk determines the time it takes to retrieve it from declarative memory: $T = e^{-A_i}$.

2 Statistical analyses

Table A1: Fixed effects of best fitting logistic mixed-effects model to fit the choice for the subject as the referent of the ambiguous pronoun in the test stories.

Formula

```
choiceSubj ~ (WMload + TopicShift + Session + TrialPosition)^4 + (1|Subject) +
(1+TopicShift|Item)
```

Predictor		Estimate	SE	z value	p-value
Intercept	***	2.123	0.348	6.11	0.000
WMload[1]		−0.207	0.250	−0.83	0.406
Topicshift[2]		0.238	0.330	0.72	0.472
Session[3]		−0.170	0.334	−0.51	0.610
Trialposition[4]	**	−0.009	0.003	−2.98	0.003
WMload:Topicshift		−0.138	0.250	−0.55	0.580
WMload:Session		−0.274	0.268	−1.02	0.307
WMload:Trialposition		0.000	0.002	0.12	0.901
Topicshift:Session		0.269	0.330	0.82	0.414
Topicshift:Trialposition		−0.003	0.003	−0.99	0.322
Session:Trialposition		−0.001	0.003	−0.24	0.810
WMload:Topicshift:Session	**	−0.676	0.250	−2.71	0.007
WMload:Topicshift:Trialposition		0.003	0.002	1.36	0.173
WMload:Session:Trialposition		0.002	0.002	0.72	0.474
Topicshift:Session:Trialposition		−0.001	0.003	−0.44	0.662
WMload:Topicshift:Session:Trial position	**	0.007	0.002	2.84	0.005

Model characteristics

[1] Contrast used for WM load: Low WM load (-1) versus High WM load (1);
[2] Contrast used for Topic shift: No topic shift (-1) versus Topic shift (1);
[3] Contrast used for Session: Session 2 (-1) versus Session 1 (1);
[4] Trial position: range 1-192; linear fit, because log-transformed predictor or more complex transformations did not improve the model.

Table A2: Fixed effects of the mixed-effects model (including all two-way interactions) to fit the reading times on the first region of Sentence 2.

Formula

```
LogRT ~ (WMload + TopicShift + TrialPosition + Session)^2 + Letters +
(1+WMload|Subject) + (1 |Item)
```

Predictor		Estimate	SE	t value[1]
Intercept	***	5.458	0.110	49.63
WMload	*	0.118	0.053	2.23
Topicshift		-0.046	0.051	-0.89
Trialposition	***	-0.005	5e-4	-9.18
Session	*	0.161	0.061	2.63
Letters		0.022	0.014	1.61
WMload:Topicshift		0.015	0.014	1.08
WMload:Trialposition	*	-0.001	5e-4	-2.52
WMload:Session		-0.047	0.064	-0.74
Topicshift:Trialposition		4e-4	5e-4	0.69
Topicshift:Session		0.012	0.028	0.41
Trialposition:Session	***	-0.002	5e-4	-4.5

Model characteristics

[1] Exact p-values are not estimated, because a random slope for WM load needed to be included for Subjects ($\chi^2(2)$=14.929, p<.001).
See Table 1 for an explanation of the contrast coding used.

Table A3: Fixed effects of the mixed-effects model (including all two-way interactions) to fit the reading times on the second region of Sentence 2.

Formula

```
LogRT ~ WMload + TopicShift + TrialPosition + Session + Letters + LogRT1 +
(1|Subject) + (1 |Item)
```

Predictor		Estimate	SE	z value	p-value
Intercept	***	3.391	0.189	17.916	0.000
WM load		0.015	0.013	1.125	>.1
Topic shift	*	0.028	0.013	2.151	0.032
Trial position	***	-0.003	6e-4	-4.659	0.000
Session		-0.040	0.030	-1.331	>.1
Letters		-0.030	0.017	-1.762	0.078
LogRT1	***	0.442	0.026	16.711	0.000

Model characteristics

See Table 1 for an explanation of the contrast coding used.

chapter 5

Context resolves ambiguity before an ambiguous pronoun is read:

An ERP study of the effects of context and working memory load on pronoun comprehension

Abstract

ERP studies on pronoun processing often elicit strong effects when the referent of the pronoun is ambiguous, an effect that has been shown to be modulated by individual differences in working memory capacity. These studies, however, typically provide only limited linguistic context. Here we examined the electrophysiological correlates of experimentally manipulated WM load during the on-line resolution of pronouns in discourse. While participants had to remember a series of either 3 or 6 digits, they read short stories introducing two agents by name, with an ambiguous pronoun in the final sentence. In half of the stories the saliency of the secondly introduced referent was increased by reversing the grammatical roles of the two referents. Interestingly, no or only small ERP effects were observed at the pronoun. Instead, clear effects of discourse structure and WM load were already found at the first introduction of the second referent. Based on these results, we argue that 1) that discourse saliency already anticipates the most likely referent of an upcoming pronoun, effectively resolving any potential ambiguity at the pronoun itself, and 2) higher WM load changes discourse processing, which may also affect off-line story comprehension. This study provides further evidence for the notion that pronoun resolution may reflect discourse ambiguity. However, as this study also suggests that disambiguation of referential roles often associated with pronoun processing might already take place at the first introduction of a second referent.

1 Introduction

Pronouns such as *he*, *she*, or *they* do not have a fixed interpretation, but refer to a character in the context that matches the pronoun in gender and number features. In many situations, more than one character in the context can be selected as the antecedent of the pronoun. As a result, listeners or readers may perceive the pronoun as ambiguous. For example, when encountering a pronoun at the end of a paragraph in a novel with many different characters, you may discover that you are not sure who the author is referring to with the pronoun. More often, luckily, readers do not seem to be aware that the pronoun can have more than one interpretation, but immediately choose the correct interpretation.

In this study we investigate whether potentially ambiguous pronouns are indeed ambiguous. Do people have to choose between possible referents when they encounter a pronoun, or is there no need for them to consider alternative referents at the pronoun because the preceding linguistic discourse context has already constrained the choice of referent? Does the structure of the linguistic discourse already invoke the necessary processes for referential disambiguation of the pronoun? As these processes obviously require an internal representation of the potential referents and their relative prominence in the linguistic discourse, cognitive aspects such as working memory capacity might affect these disambiguation processes. Therefore, a related question is whether and how pronoun interpretation depends on the amount of cognitive resources available for language processing: Can we still combine all information from the preceding linguistic context to resolve the reference of the pronoun when we are busy with another cognitively demanding task? Investigating these questions may provide insights in how listeners are able to integrate linguistic information very rapidly during language processing.

1.1 Pronominal ambiguity

Many reading time studies have reported that it takes longer to read sentences containing an ambiguous pronoun than sentences with an unambiguous pronoun with only one gender-matching referent (a.o., Caramazza, Grober, Garvey, & Yates, 1977; Ehrlich, 1980; Crawley et al., 1990). Pronouns that are ambiguous because the linguistic context contains more than one gender-matching referent, are generally interpreted as referring to the most salient character in the linguistic context, the *discourse topic*. The saliency of the characters in the linguistic context is influenced by many factors (see Arnold, 1998, for an overview), among other things, the grammatical roles of the referents in the linguistic discourse (a referent that is the subject of the previous utterance is preferred over a referent that is an object of the previous utterance), order of mention (the referent that is mentioned first is preferred over a referent that is mentioned later in the linguistic discourse), and recency (referents that were referred to in a recent utterance are preferred over referents that were referred to several utterances back), and verb causality (some verbs are biased towards one of the previously mentioned referents based on their grammatical role). However, although all these factor influence saliency, they do not always work in the same direction. Hence, it is not uncommon to encounter ambiguous pronouns.

Ambiguity of reference may not be resolved in the same way by all listeners or readers. Various studies have related differences in ambiguity resolution to individual differences in working memory capacity (a.o., Just & Carpenter, 1992; Miyake, Just, & Carpenter, 1994; Nieuwland & van Berkum, 2006; van Rij et al., 2011). For example, Nieuwland and van Berkum (2006) found that participants' scores on a reading span task correlated with the electrophysiological recordings associated with resolving the ambiguity of subject pronouns.

Interestingly, the effects of context on referent saliency and individual differences in ambiguity resolution interact. In a behavioral study (van Rij et al., 2011; van Rij, van Rijn, & Hendriks, submitted*b*), it was found that experimentally manipulating the amount of WM load during comprehension influences participants' sensitivity to the factors influencing the referent's saliency in the linguistic context, and thus affected their pronoun interpretation. In these studies, participants were asked to memorize a sequence of 3 or 6 digits for later recall before reading stories about two characters of the same gender (see Table 1 for examples). Two types of stories were tested: stories with and without a topic shift. The topic is the most salient referent in the linguistic discourse. The discourse topic is viewed here as a local notion, which can shift from one utterance to the next and is determined by grammatical prominence (cf. Grosz et al., 1995). In the stories with a topic shift the *topic* shifted halfway, as the second character replaced the first character as the grammatically most prominent referent by becoming the subject of next sentences. In all stories the final sentence started with a potentially ambiguous pronoun (*he* or *she*).

A. Story with topic shift	**B. Story without topic shift**
1. Eric gaat voetballen in de sporthal. 'Eric is going to play soccer in the sports hall.'	1. Eric gaat voetballen in de sporthal. 'Eric is going to play soccer in the sports hall.'
2. Philip vraagt Eric om mee te rijden/naar de training. 'Philip asks Eric to carpool to the training.'	2. Eric vraagt Philip om mee te rijden/naar de training. 'Eric asks Philip to carpool to the training.'
3. Philip haalt Eric na het eten met de auto op. 'Philip picks up Eric after dinner by car.'	3. Eric haalt Philip na het eten/met de auto op. 'Eric picks up Philip after dinner by car.'
4. Hij voetbalt al twintig jaar. 'He has played soccer for twenty years.'	4. Hij voetbalt al twintig jaar. 'He has played soccer for twenty years.'
Question: Wie voetbalt al twintig jaar? / 'Who has played soccer for twenty years?'	

Table 1: Example of the Dutch sentences (and the English translations) of a story with and without topic shift as used in van Rij, van Rijn, & Hendriks (2011, submitted*b*) and in this study. Time windows for ERP analyses are marked in gray.

In the 6 digit condition, which resulted in higher concurrent memory load, participants less frequently selected the subject of the previous sentence as the antecedent of the pronoun after a topic shift than in the 3 digit condition, indicating that the pronouns in the topic shift

stories became more ambiguous with higher memory load. Importantly, no effect of WM load was found on the choice of referent in the stories without a topic shift. This suggests that with higher WM load, participants were less sensitive to discourse prominence factors such as grammatical role in determining the referent of an ambiguous subject pronoun. Instead, they relied more on other factors, such as frequency and recency of mentioning. Thus, the referents' saliency during the interpretation of subject pronouns is a function of the working memory capacity that is available during sentence processing, in accordance with the findings of Nieuwland and van Berkum (2006).

In the same study (van Rij et al., 2011, submitted*b*), no differences were found between conditions in the reading times on the pronoun in Sentence 4 (see Table 1), which is surprising given earlier studies that found reading times reflecting the ambiguity of the pronoun (a.o., Caramazza et al., 1977; Ehrlich, 1980; Crawley et al., 1990). It might be that subject-paced reading times are not sensitive enough to detect the change in ambiguity in the topic shift stories, even though the preference for the first referent as the antecedent of the pronoun is reduced with 10% in the higher WM load condition compared with the lower WM condition. However, the reading times did show effects in Sentence 2. In Sentence 2, the second referent is introduced on the first word in the topic shift stories and on the third word in the stories without topic shift. On the first word, reading times were slower in the high WM load in comparison with the low WM load condition in both story types. The second word, the spillover region, showed increased reading times for stories with a topic shift. The effects of WM load and topic shift indicate that readers noticed the topic shift, and that this topic shift immediately changed the relative saliency of the referents in the discourse representation because no effects of ambiguity were found on the reading times later in the story. Thus, the introduction of the second referent in the grammatically most prominent subject position caused an immediately commitment for that referent as the topic of the story, instead of leaving the choice for a discourse topic undetermined until other processes require the identification of the discourse topic.

Given the paradoxical results discussed above, we conduct an EEG study to examine whether WM load and Topic shift interact in the on-line resolution of pronouns. More specifically, we are interested to see whether an electrophysiological analogue of the effects observed in van Rij et al. (2011, submitted*b*) can be found, indicating that the reader already commits to a particular discourse topic and thus already anticipates the most likely referent of a not yet encountered pronoun.

1.2 ERP effects of context, pronominal ambiguity and individual differences

Many ERP studies on language processing have reported a negative deflection peaking around 400 ms after stimulus onset: a N400 component (see Kutas, Van Petten, & Kluender, 2006, for a review). The amplitude of this ERP component is modulated by the semantic context and by the lexical properties of the stimulus word (a.o., Kutas & Hillyard, 1980, 1984; Holcomb & Neville, 1991). Generally, the N400 amplitude of open class words (nouns and verbs) decreases

with each subsequent word in a semantically congruent sentence, suggesting that the sentence context constraints the interpretation of the sentence (Van Petten & Kutas, 1990). Not only the sentence, but also the linguistic context preceding the sentence has been found to influence the amplitude of the N400 (a.o., St. George, Mannes, & Hoffman, 1994; van Berkum, Hagoort, & Brown, 1999; Van Petten, 1995). Van Petten (1995) reports that the N400 amplitude only decreases with word position in isolated sentences, but not when these sentences are preceded by a meaningful linguistic context. van Berkum et al. (1999) report that words that are semantically anomalous given the preceding discourse, but at the same time acceptable within the local sentence context, elicit a larger N400 than when the same words is presented in isolation without any preceding context. Thus, the N400 reflects how words fit in the preceding context.

Recently, a number of ERP studies have specifically focused on pronoun processing (a.o., Nieuwland & van Berkum, 2006; Streb, Rösler, & Hennighausen, 1999; Streb, Hennighausen, & Rösler, 2004; Swaab, Camblin, & Gordon, 2004). Streb et al. (1999) observed a larger negativity for unambiguous pronouns than for proper names in the second sentence of discourses consisting of two parallel sentences or two nonparallel sentences, in which the grammatical roles of the referents are reversed. In addition, they found that unambiguous pronouns and proper names in the nonparallel discourse structures also elicited a larger negativity than in the parallel discourse structures. This finding is in line with the results from the behavioral study (van Rij et al., 2011, submitted*b*) since a change in grammatical role influenced the resolution processes.

Using a different manipulation of the saliency of the potential references, Nieuwland and van Berkum (2006) found that ERPs can reflect contextual influences on pronoun resolution. They examined whether the ERP signal reflects influences of world-knowledge biases on pronoun resolution by comparing sentences with an ambiguous pronoun (two potential referents) with sentences with an unambiguous pronoun (one potential referent). The ambiguous pronoun sentences were manipulated so that different levels of ambiguity could be compared.

The sentences with an ambiguous pronoun that was weakly biased towards one of the two referents elicited an increased and sustained frontal negative shift starting around 300 ms (e.g.,*The chemist hit the historian, while he...* in 53% the historian is selected as antecedent; Nieuwland & van Berkum, 2006), compared to the sentences with an unambiguous pronoun. In contrast, the sentences with a pronoun that was moderately biased towards one of the referents did not differ from the disambiguated sentences (e.g. *Linda invited Anna, when her...* in 70% Linda is selected as antecedent). Interestingly, the distribution of responses, which reflects how ambiguous a pronoun was, was similar for the moderately biased sentences used by Nieuwland and Van Berkum and the stimuli used the study of van Rij et al. (2011, submitted*b*). In our study, no effects were found in the readings times on the ambiguous pronoun, but effects of the manipulation of saliency were found earlier in the discourse. This observation raises the question whether similar signatures of disambiguation can be found in ERP components elicited by the shift in discourse topic.

Nieuwland and Van Berkum also observed that only participants with a higher working memory (WM) capacity score showed a sustained negative shift for the sentences with am-

biguous pronouns (collapsed over moderate and low bias conditions) relative to sentences with unambiguous pronouns. In contrast, the participants with a low WM capacity score did not show a difference. They refer to this slow negative shift as *Nref* or referential ambiguity effect (van Berkum, 2008) because it is different from the N400 in duration (similar onset latency to N400, but sustaining up to more than a second) and distribution (more anterior than the N400). Interestingly, a similar negative shift is also associated with working memory demands (a.o., King & Kutas, 1995; Münte, Schiltz, & Kutas, 1998; Phillips, Kazanina, & Abada, 2005) . For example, King and Kutas (1995) showed that high WM capacity readers showed a larger difference in the slow frontal negative shift when comparing the reading of subject relative sentences to object relative sentences. These slow negative shifts are not specific for language processing, but are also observed in non-verbal memory and attention tasks (see Bosch, Mecklinger, & Friederici, 2001; Kutas & Federmeier, 2007; van Berkum, Koorneef, Otten, & Nieuwland, 2007, for reviews).

1.3 Current study

In the current study, we address the question whether people resolve pronominal ambiguity in the discourse by an early commitment for one of the referents as the discourse topic instead of resolving the ambiguity when the pronoun is encountered. As previous research indicates that the amount of WM capacity available for language processing may change the sensitivity for the ambiguity in the discourse, we also address how WM load affects on-line story comprehension. Therefore, we measure EEG while participants read short stories similar to the stories used in van Rij et al. (2011, submitted*b*), as shown in Table 1. We manipulate memory load by a secondary task that elicits higher or lower WM load. By recording the EEG signal during reading we want to examine 1) whether a potential ambiguity in the discourse is resolved immediately after a topic shift, or only when reading a pronoun and 2) whether higher WM load changes discourse processing, which may affect off-line story comprehension as well. Based on the reviewed literature, we focus on ERP components in the N400 region and slow negative shifts, which may indicate high processing load.

2 Experimental materials and methods

The current study uses a dual-task set-up to investigate the influences of topic shift and WM load on on-line pronoun processing. The design and materials were adapted from van Rij et al. (2011, submitted*b*).

2.1 Methods

2.1.1 Digit task

At the start of each trial, participants had to memorize a sequence of either 3 digits (low WM load condition) or 6 digits (high WM load condition). Digits were shown for 1 second each in the center of a computer screen. The digits were pseudo-randomly chosen from 1 to 9 to

ensure that not all digits were the same. After completing a trial of the story-reading task, the participants had to recall the memorized digits. Feedback was provided for the digit task, after the trial was completed.

2.1.2 Story-reading task

After the presentation of the digits, participants read stories consisting of four sentences each (see Table 1), followed by a comprehension question. The stories were presented word by word with a variable serial visual presentation procedure (cf. Nieuwland & van Berkum, 2006), which combines a relatively natural temporal reading pattern with fixed presentation durations for critical words. The presentation duration for non-critical words was dependent on their length, with 30 ms per letter (with an upper bound of 8 letters / 240 ms) added to an offset of 190 ms. The critical words (the first four words of sentence 2 and the first two words of sentence 4, highlighted in Table 1) were presented with a fixed duration of 325 ms (based on the mean number of letters of the critical words). Sentence final words stayed 300 ms longer on the screen. Between words, an empty screen was presented for 150 ms, and between sentences, for 1000 ms. Importantly, the first name of each story was presented with a duration of 1000 ms because reading the first name was crucial for detecting the topic shift. Subjective evaluations indicated that the resulting temporal pattern allowed for a relatively natural reading pattern. We recorded the real onset and offset times for each presented word to be able to align the EEG data to the precise onset of the target words. As a result of small delays caused by, among other things, the refresh rate of the monitor, the critical words followed each other after on average 490.2 ms (sd=1.4), the duration of the empty screen following each word included.

After reading the four-sentence story, a question was presented with two answer alternatives presented on the left and right side of the screen. To answer the question, participants had to press either the left arrow key on the keyboard (for the left answer) or the right arrow key. After answering the question, they had to type in the digits that were presented at the beginning of the trial. At the end of each trial, participants received feedback on the digit task to ensure sufficient focus on the WM task. Before the start of a new trial, participants were instructed to remove their hands from the keyboard.

2.1.3 Design

Stories. Every story featured two characters of the same gender. The final sentence of each story started with an ambiguous subject pronoun *hij* ('he') or *zij* ('she'): the pronoun could refer to both characters, so the only clue to the interpretation of the pronouns was the structure of the story. We designed two variants of every story (see Table 1), in which we manipulated whether there was a topic shift or not. The topic shift is realized by placing the second character (*Philip*) in subject position in Sentence 2. Thus, the only difference between the two variants of each story was the grammatical role of the referents, causing a difference in relative prominence. In the no shift stories, the first character was the subject of all sentences before the pronoun. In the topic shift stories, the second character became the subject

of the second and third sentence. Subjective evaluations indicated that both variants of the stories were equally natural and that world-knowledge biases did not affect the interpretation of the pronoun. When there was no topic shift, we expected participants to prefer the firstly introduced character as the referent of the ambiguous pronoun because this referent was the grammatical subject in the sentence preceding the pronoun. In addition, it was firstly introduced, and most frequently mentioned. When there was a topic shift, we expected participants to prefer the second character because this referent was the grammatical subject in the sentence preceding the pronoun. At the end of every experimental story a question was presented to elicit the preferred interpretation of the ambiguous pronoun.
Lists. In total, 160 stories were designed in two different variants, with the variants randomly assigned to one of two lists. Of the 160 stories, 64 stories were followed by a question regarding the preferred interpretation of the ambiguous pronoun. The other 96 stories were followed by a question about the first or second sentence of the story, to discourage reading strategies.

2.1.4 Procedure

Participants were randomly assigned to one of the two lists. The experiment consisted of two blocks: a low WM load block (3 digits) and a high WM load block (6 digits). The order of blocks was counterbalanced; within blocks the items were randomly distributed. Participants received separate instructions for each block. The experiment started with a practice trial suited for the current WM load condition.

2.2 Participants

Twenty-one undergraduate students (5 men; 2 left handed[1]; age range 17-24[2], mean age 19) participated in the experiment in exchange for course credits or money. Seven participants did not finish all trials in the experiment because of technical problems. All participants were included for statistical analyses with on average 146 trials per participant (range 80-160), and appropriate care was taken during analysis to address unbalanced numbers of observations in the different cells.

2.3 EEG recordings

EEG was recorded at 500 Hz from 30 electrodes fitted in an elastic cap (AFz, Fz, FCz, Cz, CPz, Pz, FP1/2, F7/8, F3/4, FC5/6, FC1/2, T7/8, C3/4, CP5/6, CP1/2, P7/8, P3/4, O1/2). Horizontal eye movements were monitored with a bipolar electrode pair, positioned left from the left eye, and right from the right eye. Vertical eye movements and blinks were recorded with a bipolar electrode pair, under and above the left eye. FieldTrip (version 2011-12-08; Oostenveld, Fries, Maris, & Schoffelen, 2011), a Matlab software toolbox for EEG analysis, was used for preprocessing and artifact rejection: The data of the second and fourth sentences was re-referenced to the mean of left and right mastoids, time locked to the onset of the first word of that sentence, segmented (500 ms before to 2500 ms after the sentence onset), filtered (Butterworth bandpass filter 0.1-125Hz, and band stop filter 48-52Hz), and down sampled to 256

Hz after. All trials were semi-automatically screened for muscle artifacts, alpha waves and recording problems. In our artifact rejection procedure, 7% from the recordings of Sentence 2, and 8% from the recordings of Sentence 4 were excluded. Eye movements, blinks, heart activity, and slow voltage shifts were removed from the data using Independent Component Analysis (ICA; using the Runica algorithm in FieldTrip). After artifact rejection, the data was filtered (band-pass filter at 0.1-20Hz), baselined per trial per participant (50 ms before to 50 ms after the onset of the sentence) and down sampled to 75Hz for analysis.

3 Results

In this section, we first discuss the performance on the digit task and the off-line story comprehension results, i.e., answers on the questions and the related reaction times, before presenting the ERP data.

3.1 Performance digit task

In the low WM load condition, participants remembered 78.0% of the digit sequences correctly, but in the high WM load condition this percentage was significantly reduced to 59.5% (paired-t(19)=-6.89; p<.001). We did not find any effect of story condition on the percentages correct trials in the digit task. The performance on the digit task was used as a control measure, since the WM load manipulation depends on participants memorizing the digits while reading the stories. Therefore, all trials with too many mistakes in the memorized digits sequence were excluded from further analyses: in the low WM load, only correct trials were included (78.0% were included; 328 of 1475 trials were removed), but as participants are more likely to make errors during recall when a longer list of digits has to be recalled even if the list was memorized, one mistake was allowed in the sequence of six digits in the high WM load condition (80.9% were included; 296 of 1600 trials were removed).

3.2 Performance comprehension questions

After excluding the trials in which the digit sequence was not memorized correctly, we analyzed the answers on the 64 story comprehension questions eliciting the preferred interpretation of the ambiguous pronoun. Figure 1 (right panel) shows the proportions of selection of the subject of the previous sentence as the antecedent of the pronoun (in the example in Table 1, this would be *Philip* in stories with topic shift, and *Eric* in stories without topic shift). Participants clearly make use of the structure of the discourse because in around 70% of the trials the subject of the previous sentence is selected as the antecedent of the pronoun. This preference of 70% for one of the referents is comparable with the moderately biased pronouns in the Nieuwland and van Berkum (2006) study, for which they did not measure a difference in ERPs with the unambiguous pronouns.

We examined the effects of the two level factors *Topic shift* and *WM load*, and *Trial*, the position of the trial in the experiment, on the choice for the subject of the previous sentence (yes

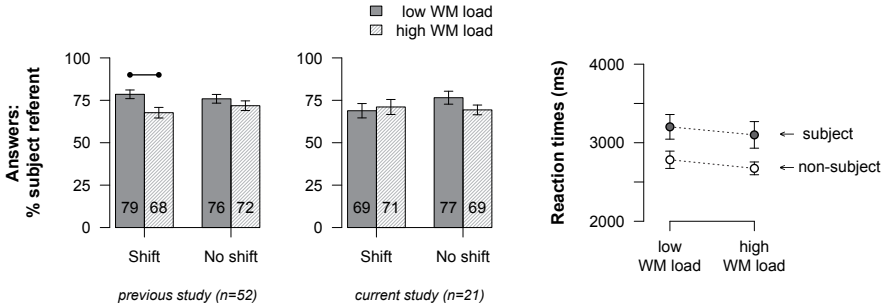


Figure 1: Preference for the subject as referent of the pronoun in stories with and without topic shift (±1SE) and reaction times (±1SE). *Left:* results from previous behavioral study using the same stories (adapted from Van Rij et al., submitted). *Center:* results from current study. *Right:* reaction times from current study.

or no) using logistic mixed-effects models (cf. Baayen, 2008). More complex models that included additional predictors did not show qualitatively different effects. Participant and item (i.e., both variants of a story were labeled as the same item) were included as crossed-random effects. Using a stepwise variable deletion procedure, starting with the complete interaction model, we found only a marginal significant main effect for *WM load* ($\chi^2(1)= 3.68$; p=.055), suggesting that with higher WM load the subject of the previous sentence is less often selected as the antecedent of the pronoun. The lack of effects in the answers on the comprehension questions is not surprising, given that the effect of WM load on the use of discourse structure is very subtle. Although we tried to make the stories as neutral as possible towards both referents with respect to implicit verb biases, the effect of discourse structure may be masked by idiosyncratic preferences of the words in the sentence. Also, the smaller number of participants in this EEG study compared to the previous behavioral experiment might have reduced power.

Additionally, we analyzed the reaction times on the story comprehension questions (see Figure 1, right panel). Mixed-effects models were used to analyze the logarithmically transformed reaction times with *Topic shift*, *WM load*, and *Trial* as predictors, but also *SubjectAnswer* (whether or not participants selected the previous subject as the antecedent of the pronoun). Using a stepwise variable deletion procedure, starting with the complete interaction model, we found a significant main effect for *SubjectAnswer* ($\chi^2(1)= 21.74$; p<.001), suggesting that participants are slower to select the answer when they do not choose the subject of the previous sentence as the antecedent of the pronoun. In addition, a significant interaction was found between *WM load* and *Trial* ($\chi^2(1)= 21.23$; p<.001). This interaction suggests that as the experiment progresses, participants become faster, but the reaction times increase more in the high WM load condition, although participants are generally slower in the high WM condition than in the low WM load condition. Thus, a significant effect of WM load is present in the reaction times, and a similar effect is marginal significant in the answers.

The main goal of the current experiment was collecting electrophysiological markers of disambiguation. Since both filler and test stories have the same discourse structure, and only differ in the comprehension questions, all stories were used for EEG analyses.

3.3 EEG data

Figure 2 shows the grand averages for the Cz electrode of each condition in Sentence 2 and Sentence 4 of the stories. For ease of interpretation, we will focus our attention on the Cz electrode, which reflects N400 differences (a.o., Münte, Heinze, & Mangun, 1993). The same three regions of interest that were highlighted in Table 1 are indicated with gray in Figure 2: the first and third word of Sentence 2 (the two proper names), and the first word of Sentence 4 (the pronoun). Sentence 2 initiates a topic shift in the topic shift stories by introducing the second referent in subject position. In the stories without a topic shift, the second referent is introduced with the third word in the grammatically less prominent object position. Sentence 4 always starts with a pronoun, that can in principle refer to either of these referents.

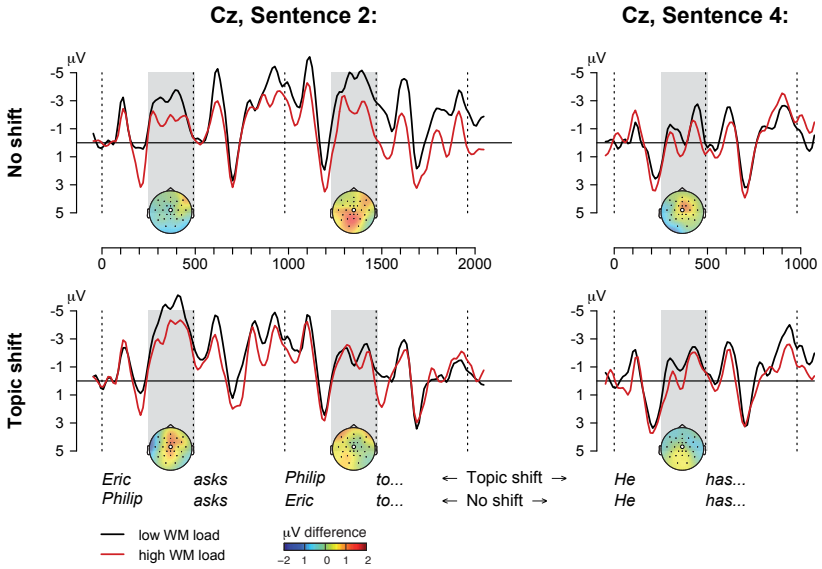


Figure 2: Grand average at electrode Cz elicited when reading word 1 to 4 in Sentence 2 (left panel) and word 1 and 2 in Sentence 4 (right panel). The first row presents the stories without topic shift; the second row presents the stories with topic shift. The black lines indicate the low WM load condition, and the red lines indicate the high WM load condition. The gray areas highlight the N400 regions (250-500 ms) following the presentation of a referring expression (proper name or pronoun). The scalp distributions present a difference topography calculated over these regions (high WM load minus low WM load).

The first row in Figure 2 shows the ERPs for stories *without* a topic shift. In Sentence 2 (left panel), the low WM condition shows a slow sustained negative shift relative to the high WM condition. This negative shift is not present in Sentence 4 (right panel). The second row in Figure 2 shows the ERPs for stories *with* a topic shift. Sentence 2 (left panel) shows a positive shift with respect to the baseline (onset sentence), and a decrease in amplitude when the sentence unfolds. The ERPs of the topic shift stories in Sentence 4 (right panel) are visually similar to the ERPs of the stories without topic shift (top row, right panel), with also no large differences between the WM load conditions.

3.4 Analyses

In this study, we use generative additive models (GAMs, Wood, 2006; see Tremblay & Baayen, 2010, for an application of GAMs to EEG data from a immediate free recall experiment) to analyze ERPs because this approach is designed to capture differences over time within a single analysis. GAMs are generalized linear models in which linear predictors can be specified by non-linear functions of predictor variables (i.e., *smooth functions*). These functions are estimated using a penalized regression method called penalized iteratively re-weighted least squares, and the parameters for each smooth function such as the degree of smoothness are estimated using cross validation (see Wood, 2006, chapters 3 and 4). These estimation processes that determine the smooth functions and parameters are designed to avoid overgeneralization and overfitting of the data.

We use GAMs to estimate the effects of topic shift and WM load on the time course of the un-averaged, time locked EEG data per sentence per participant. We did not use participant averages for the analyses, because of the different numbers of items per cell. Three time windows were defined for analyses based on the three regions of interest. The first time window starts 50 ms before the onset of the first word of Sentence 2 and ends 50 ms after the onset of the third word of the same sentence. The second time window starts 50 ms before the onset of the third word of Sentence 2 and ends 50 ms after the onset of the fifth word. The third time window starts 50 ms before the onset of the first word of Sentence 4 and ends 50 ms after the onset of the third word of the same sentence, thus including the pronoun. The time windows are 1080 ms long and included the presentations of two words, so that also later ERP differences could be detected.

For each time window, the same analysis procedure was applied to all electrodes. Here we focus on the Cz electrode for explaining the analyses and interpretation in more detail, followed by an evaluation of the distribution of the effects based on the analyses of all electrodes. In each analysis, we first removed the main trends in the EEG data over time, followed by the time trends of individual subjects, and the random intercepts of individual items. To investigate whether the residual data is influenced by *Topic shift* and *WM load* over time, we included smooth functions to account for the differences in time trends for these factors and the interaction between *Topic shift* and *WM load*. In addition, we included a predictor for *Trial*, to account for the variability in the time trends due to order of trials in the experiment. We compared different models using a stepwise variable deletion procedure, starting with the

complete interaction model. Below we first discuss the analysis of the three time windows for the Cz electrode because this electrode can reflect N400 differences (a.o., Münte et al., 1993). In Section 3.4.4, we will briefly discuss the analyses for the other electrodes.

3.4.1 Analysis Sentence 2, word 1 and 2 for Cz

In line with the leftmost panels of Figure 2, which suggest an influence of *Topic shift* and *WM load*, the best-fitting model of the Cz data for the first time window includes the main effects of Topic shift ($F(9.76, 184812)=21.37$; $p<.001$) and *WM load* ($F(4.97, 184812)=35.97$; $p<.001$). Adding the interaction of these two factors did not significantly improve the model (see Appendix 1, Table 1). To appreciate the contributions of the individual factors, Figure 3 depicts the estimated effects of the factors and the main effect of time.

The first row of Figure 3 graphs the main trend in the ERP data over time that was removed before testing the effects of *Topic shift* and *WM load*. This main trend captures all the typical ERP components with are shared by all four conditions, such as the N1-P2 complex (100 and 200 ms after the onset of the word, indicated with a dotted vertical line) followed by a N400, peaking around 400 ms after word onset. Since we are interested in the relative effects of the two main predictors, we will focus on the additive effects of *Topic shift* and *WM load*. Hereto, we fitted GAMs to the residual data of the main trend over time, but also took out the variability associated with participants, items and trial position (by removing the estimated effects of these factors). The second row in Figure 3 shows the estimated effects of *Topic shift* and *WM load* on the residual data. The gray dashed line indicates the effect of *Topic shift*. Because contrast coding was used, all predictors are estimated with respect to the baseline condition: the stories without topic shift in the low WM load condition, indicated by the horizontal line through zero. Therefore, the gray dashed line shows the estimated difference of stories with a topic shift compared to stories *without* a topic shift. Stories with a topic shift elicit an increased negativity starting as early as 150 ms after the onset of the first word and continuing until 700 ms with a peak at at 400 ms. The red solid line shows the effect of WM load, namely the estimated difference of the high WM load condition relative to the low WM condition. In the high WM load condition the ERPs are more positive, starting around 200 ms. After a positive peak at 350 ms, the high WM load condition stays more positive than the low WM load condition.

In the third row of Figure 3, the estimated effects of Topic shift and WM load are added to the main trend over time. During processing the first word (the first proper name), the two main effects result in a larger negativity for stories with a topic shift, and a larger negativity for the low WM load conditions. The effect of *Topic shift* is larger than the effect of *WM load*, and both topic shift stories have a larger amplitude than the stories without topic shift. However, the pattern changes during processing of the second word (the verb): the negative effect of Topic shift changes into a small positive effect. After reading the second word, stories *without* a topic shift show a larger negativity than stories *with* a topic shift, just as the low WM load conditions compared with the high WM conditions.

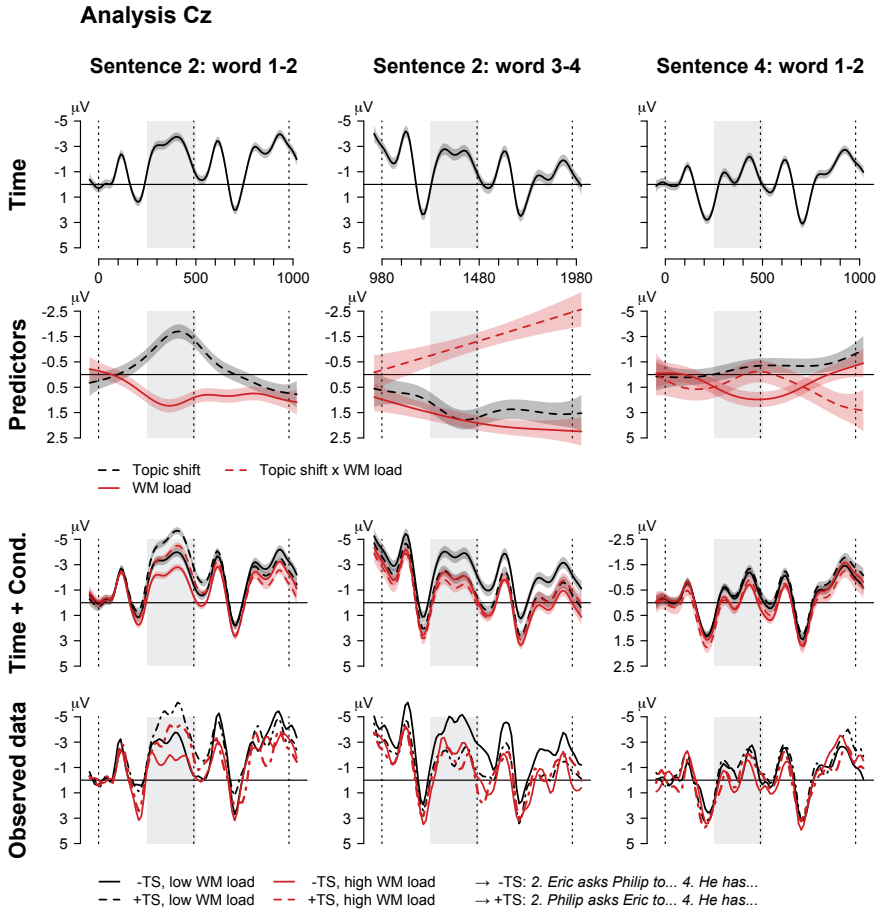


Figure 3: Estimated effects (±1SE) of the predictors on the ERP at Cz over time in three time windows, each starting on the onset of the referring expression (note that the regions are partly overlapping). The dotted vertical lines indicate the onset of a new word, and the shading around the lines in rows 1-3 represent ±1SE. First row: the main effect of Time. Second row: the additive effects of Topic shift (gray, dashed line), WM load (red solid line) and Topic shift x WM load (red, dashed line). Note that the interaction was not significant in the first time window. Third row: the effects of predictors added to the main effects over time for each condition. Fourth row: the observed data for the Cz electrode (see also Figure 2 for a different presentation of the same data).

These results are similar to the reading times in our previous study (van Rij et al., 2011, submitted*b*), in which we found main effects of WM load and Topic shift on the first two words of Sentence 2. The higher amplitude for reading a new name in the topic shift stories than reading the name of an earlier introduced referent in the stories without topic shift is in line with earlier studies (Streb et al., 1999; Swaab et al., 2004) and may reflect the additional processing required for the integration of new information in the discourse representation. The lower N400 amplitudes in the high WM load conditions may indicate a smaller effect of context on reading referring expressions due to less WM capacity available for discourse processing, as the N400 is assumed to reflect how words fit in the preceding context.

3.4.2 Analysis Sentence 2, word 3 and 4 for Cz

The best-fitting model of the Cz data for the second time window included the main effects of *Topic shift* ($F(8.23, 184816)=18.07$; $p<.001$) and *WM load* ($F(3.98, 184816)=62.77$; $p<.001$), but also the interaction *Topic shift x WM load* ($F(2.12, 184816)=38.83$; $p<.001$) (see Appendix 1, Table 2).

The middle panel of Figure 3 shows the estimated effects for the second time window. The first row displays the main trend over time while participants process word three and four. The second row plots the estimated effects of *Topic shift* (dashed gray line), *WM load* (solid red line) and the interaction *Topic shift x WM load* (dashed red line). *Topic shift* and *WM load* elicit a positive shift, indicating that stories with a topic shift become more positive than stories without a topic shift during this time window, and that ERPs in the higher WM load condition become more positive during this time window than the ERPs in the low WM load condition. The negative estimated effect of the interaction indicates that the stories with topic shift in the high WM load condition do not become extra positive, because the increasing negativity of the interaction effect reduces the increasing positive effects of topic shift and WM load. Together, these effects indicate that all conditions become increasingly more positive over time, except for the baseline condition: the amplitudes in the stories without topic shift in the low WM condition remain the same, while the other conditions become relatively more positive. This is illustrated in the third row of Figure 3, where the estimated effects of *Topic shift*, *WM load*, and *Topic shift x WM load* are added to the main trend over time. The positive shift for the high WM load conditions relative to the low WM load conditions starts very early after the onset of the sentence. The positive shift for the stories with a topic shift starts around 800 ms, during the processing of the verb.

Interestingly, during the processing of the second name in the sentence (word 3) the positivity peaks around 450 ms after the onset of that name (i.e., 1430 ms after sentence onset), indicating a reduced N400 amplitude. The effects of *WM load* and the interaction *Topic shift x WM load* do not show a similar peak in the second time window, suggesting a more linear trend over time. These effects indicate that only topic shift influences the N400 amplitude, but not WM load. In general, the slow positive shift of *Topic shift* and *WM load* that are found over the whole time window, and the slow negative shift of the interaction predictor reflect a difference between the baseline condition and the other conditions: the baseline remains negative over time, whereas the other conditions become increasingly more positive as the

sentence unfolds. As explained before, sustained negative shifts may reflect increased memory demands or processing load. This implicates that the baseline condition in Sentence 2 may require more memory demands than the other conditions.

3.4.3 Analysis Sentence 4, word 1 and 2 for Cz

The best-fitting model of the Cz data for the third time window included the main effects of *Topic shift* (F(3.77, 182299)=5.45; p<.001), *WM load* (F(8.29, 182299)=8.08; p<.001), and also the interaction *Topic shift x WM load* (F(5.98, 182299)=6.67; p<.001) (see Appendix 1, Table 3). The rightmost panel of Figure 2 shows the estimated effects of the main effects and interaction during the pronoun sentence (Sentence 4).

The data of Sentence 4 is baselined to the onset of that sentence. The first row shows the main trend over time. Reading a pronoun elicits a slightly reduced N400 in comparison with reading the proper names in Sentence 2 (as can be seen by comparing the gray areas of Sentence 2 with the gray areas of Sentence 4 in the first row of Figure 3). The N400 consists of two smaller peaks around 300 and 450 ms. The second row of Figure 3 shows the estimated effects for *Topic shift*, *WM load*, and *Topic shift x WM load*. Although the main effect of Topic shift was statistically significant, inspection of the plot indicates that the effect size is limited since the line reflecting the estimated effect hardly deviates from zero. WM load elicits a positive effect peaking around 500 ms after pronoun onset. The interaction shows the reversed pattern, indicating that the effect of WM load is reduced for the stories with topic shift. Thus, in the high WM load condition, the N400 amplitude is lower than in the low WM load condition after reading the pronoun, and this effect is largest for the stories without topic shift. The third row of Figure 3 shows this interaction. Note that the differences in Sentence 4, although significant, are much smaller than the effects found in Sentence 3.

3.4.4 Distribution of effects

The analyses of Cz show differences in N400 amplitude on the referring expressions in Sentence 2 and 4, with the largest effect size for topic shift. In addition, WM load shows more slow wave effects in Sentence 2, and a smaller and shorter effect in Sentence 4. To test whether the slow effects of WM load are similar to the slow negative shift elicited by referential ambiguity (Nieuwland & van Berkum, 2006) or processing load (a.o., King & Kutas, 1995), we analyzed all electrodes using the same protocol as described for Cz. The slow negative shift is generally distributed over the fontal part of the scalp, slightly left lateralized (see Bosch et al., 2001; Nieuwland & van Berkum, 2006; King & Kutas, 1995). The N400 elicited by context effects is found in centroparietal regions of the scalp. Therefore, we would expect to see this distribution for the effects of *Topic shift*.

Figure 4 indicates for each time window where the main effects of *Topic shift* and *WM load* and the interaction *Topic shift x WM load* are significant and needed to be included in the model. The color coding indicates the level of significance and the direction of the effect: the red colors indicate that the effect results in a positive shift (averaged over the time window), and the blue colors indicate that the effect results in a negative shift. Note that we did not

need to include the interaction between Topic shift and WM load in the analysis model for Cz, but this interaction does significantly contribute to the explanation of the data in the lateral and posterior electrodes.

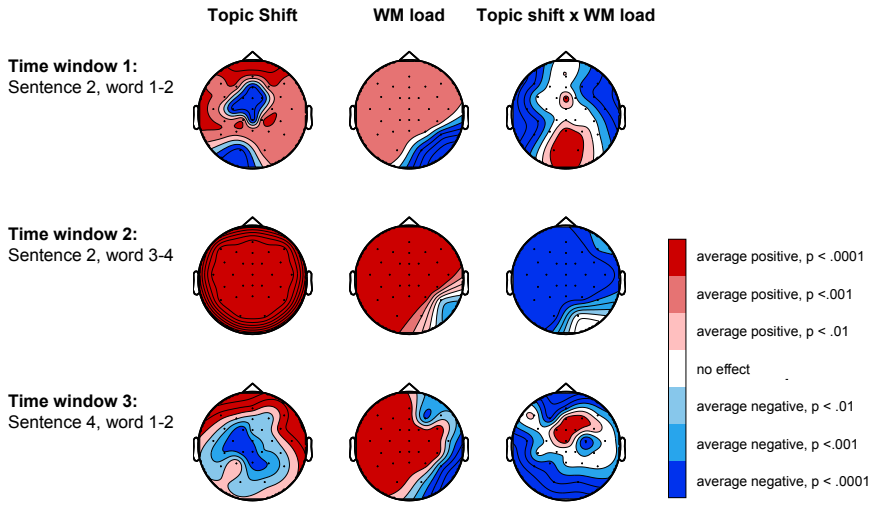


Figure 4: Distribution of the effects of Topic shift, WM load and the interaction Topic shift x WM load for each time window (viewed from above). White areas indicate that the respective predictor is non-significantly contributing to the explanation of the variance of the statistical model. The red colors indicate that the predictor shifts the ERP signal in a positive direction; the blue colors indicate that the predictor shifts the ERP signal in a negative direction.

The leftmost panel of Figure 4 shows the distribution of the predictor *Topic shift*. The central dark blue area on the scalp in time window 1 (top row) indicates that in this time window the stories with topic shift are significantly more negative than the stories without topic shift. The distribution of this negativity clearly indicates a central N400 effect. Time window 3 (bottom row) also shows a negative effect, localized centroparietal, but this effects is more broadly distributed and less clear, reflected by the lighter colors that indicate a lower level of significance. This may be due to the difference in effect size (which is not graphed in this plot, but see Appendix 2, Figure 1): in Sentence 2, the maximum estimated difference between stories without and with topic shift is -1.62 μV (estimated for FCz, from 250-500 ms after sentence onset), whereas in Sentence 4 the maximum estimated difference between stories without and with topic shift is -0.39 μV (estimated for P4, from 250-500 ms after sentence onset). In contrast to time windows 1 and 3, time window 2 shows a general positivity over the whole scalp, as we also found in Cz (previous section). To summarize, the left panel of Figure 3 shows that topic shift stories elicit a clear N400 difference with stories *without* topic shift at the start of both sentences.

The middle panel of Figure 4 graphs the effect of WM load. The red color in all time win-

dows indicates a significant positive shift distributed over the fontal part of the scalp, slightly left lateralized. The localization of this positive shift is similar to the slow ERP wave reported in earlier studies. Interestingly, the effect reaches a higher level of significance in time windows 2 and 3, indicated by the darker red color. Thus, in the higher WM load condition the ERP signal is generally more positive than in the lower WM load condition. The estimated effects are considerably larger in Sentence 2 than in Sentence 4: in Sentence 2, the maximum estimated difference between stories without and with topic shift is 1.47 μV (estimated for F8, from 250-500 ms after sentence onset), whereas in Sentence 4 the maximum estimated difference between stories without and with topic shift is 0.82 μV (estimated for Cz, from 250-500 ms after sentence onset).

The right panel of Figure 3 shows the interaction effect of Topic shift x WM load. The clearest pattern is found in time window 2 (middle row), which is very similar to the distribution of WM load (middle panel of Figure 3), although in the oppositive direction. In time window 2, Topic shift and WM load cause a positive shift, but the interaction reduces this positive shift for stories with topic shift in the high WM load condition. Together with the analyses of Cz, this pattern suggests that the ERP of the baseline condition (stories without topic shift in the low WM load condition) is generally more negative than the other conditions (see Appendix 2 for the estimated scalp distributions for the different conditions). In time window 1 a similar interaction is visible, but only lateral. In the central region the interaction is not significant, which may be caused by the N400 effects in the same time window. The lateral interaction effects may indicate the start of the general positive shift in comparison with the baseline condition that is visible in time window 2.

To summarize, the effects of Topic shift differ per time window: a topic shift causes a large negative amplitude at the start of Sentence 2, when a new referent is introduced in prominent position. A similar negativity is found at the start of Sentence 4, when the ambiguous pronoun is processed, but the difference between the topic shift conditions is very small. The observed differences here are mostly due to WM load. Later in Sentence 2 a positive deflection follows in the stories with a topic shift compared to the stories without topic shift. The more positive effect of WM load is consistently found in all time windows, although the size of the effect is reduced in the final sentence of the stories, and is localized in the left and frontal electrodes. The distribution of the effect of Topic shift suggests an N400 difference, and the distribution of the WM load effects are similar to the earlier reported slow wave components that are associated with memory demands and processing load.

4 Discussion

In this study we aimed to investigate how linguistic context and WM load influence the processing of potentially ambiguous pronouns. We performed a dual-task experiment in which we manipulated WM load during story comprehension. The final sentence of each story started with an ambiguous pronoun. In half of the stories, the grammatical roles of the two referents in the stories were reversed in the second and third sentence, signaling a shift in discourse topic. To examine whether the WM load caused by the secondary task and the dis-

course structure of the story influenced pronoun processing, we measured EEG while participants read the short stories. We analyzed the ERP data from the second sentence, where the second proper name was mentioned in a less prominent object position (no shift stories) or in the more prominent subject position (topic shift stories). We also analyzed the ERP data of the final sentence that started with a pronoun. Analysis of the ERP data reveals large effects in Sentence 2, but much smaller effects on the pronoun in Sentence 4.

4.1 Pronominal ambiguity

Our first question was whether ambiguous pronouns are indeed ambiguous or whether the structure of the discourse already resolves the ambiguity of the pronoun. If the potential ambiguity is not resolved before the pronoun is encountered, and the pronoun is ambiguous, then we would expect to find differences in the N400 region caused by the discourse structure and WM load, similar to previous research (a.o., Nieuwland & van Berkum, 2006). Indeed, we do find significant effects of topic shift, WM load and the interaction between topic shift and WM load starting around 200 ms after pronoun onset: higher WM load reduced the amplitude of the N400 on the pronoun, but on the verb only the stories with topic shift in the high WM load condition elicited a lower N400. The effects in the final sentence were significant, but relatively small. The effect of topic shift is hardly visible in Figure 2. Therefore, we are hesitant to conclude that these pronouns are ambiguous. These small effects contrast with earlier reading time studies (a.o., Caramazza et al., 1977; Ehrlich, 1980; Crawley et al., 1990) that showed increased reading times for pronouns with two gender-matching antecedents compared to reading times for pronouns with only one gender-matching antecedent present in the discourse. On the other hand, the small effects on the pronoun are in line with the findings of Nieuwland and van Berkum (2006), who report no significant differences between unambiguous pronouns and moderately biased pronouns. Therefore, one explanation for the small on-line effects could be that our stories were not sufficiently ambiguous to elicit a clear ERP effect of pronoun ambiguity. However, our behavioral results and reaction time data indicate that WM load causes a change in ambiguity, which we would expect to be reflected in the on-line pronoun data. Thus, our results are in the same direction as the results of van Rij et al. (2011, submitted*b*), who report a significant change in ambiguity due to WM load, but no effect at all in the reading times on the pronoun. In the current study we find only small effects of WM load and topic shift on reading the pronoun, although the off-line answers suggest that pronoun interpretation is affected by WM load.

4.2 Discourse resolves ambiguity

Another explanation of the small effects of pronominal ambiguity on reading the pronoun relates to our second question, whether the structure of the linguistic discourse already has resolved the potential ambiguity before the pronoun is encountered. If the alternative referents are not considered at the pronoun because the preceding discourse has already constrained the choice for a referent, then we would expect to find differences in the N400 region caused by topic shift and WM load already early in the discourse. Note that the earlier mentioned al-

ternative explanation, that topic shift and WM load do not cause a difference in ambiguity at all, would be supported if we did not find any differences in the second sentence or fourth sentence. However, this is not the case: On the first proper name in Sentence 2 significant main effects of topic shift and WM load were found that modulated the N400 amplitude. Interpreting a new referent in subject position caused a larger negative deflection, whereas higher WM load caused a positive deflection around the start of the N400. These effects in Sentence 2 are much larger in amplitude than the effects found on the pronoun in Sentence 4.

The reducing effect of WM load on the N400 elicited by the referring expressions reflects a smaller influence of the preceding discourse on the processing of referring expressions, as the N400 is assumed to reflect how the word fits in the preceding context (a.o., van Berkum, 2008; van Berkum et al., 1999; Van Petten & Kutas, 1990; Van Petten, 1995). This effect is in line with van Rij et al. (2011, submitted*b*, submitted*a*), who propose on the basis of computational simulations that the constraining effect of discourse on sentence processing may be reduced when listeners do not have sufficient WM capacity for using relevant information from the discourse, such as information about grammatical role.

To summarize, the effects of WM load and topic shift are considerable larger in Sentence 2 than on the ambiguous pronoun in Sentence 4. Therefore, we argue that the discourse structure largely resolved the potential ambiguity before the pronoun is encountered.

4.3 The cost of referent introduction

The larger amplitude on the *first* proper name of Sentence 2 for stories with a topic shift relative to stories without a topic shift may result from the difference between a new name, which requires the integration of new information in the discourse representation, and repeating the name of an already introduced referent, which only requires retrieving information associated with this referent from the discourse representation (Swaab et al., 2004). Streb et al. (1999) suggest that the integration of information in an existing discourse representation may require more processing demands if the referents do not have the same grammatical position as in the sentence before. The ERP effect of topic shift found on the *second* proper name in Sentence 2 in the current study supports the idea that introducing a new referent requires additional processing effort: On the second proper name, the N400 amplitude is higher in the stories without topic shift, where a new referent is introduced in this position, than in the stories with topic shift, where the firstly introduced referent is repeated in this position.

4.4 Effects of limited WM capacity

The N400 effects of topic shift and WM load were followed by a general increasing slow positive shift, starting around the onset of the second proper name. As mentioned before, the reduced N400 in the high WM load condition may reflect a reduced constraining influence of the preceding discourse on the interpretation of a sentence. In addition, a significant interaction between WM load and Topic shift was found later in the sentence (but earlier for lateral electrodes, see in Figure 3). The stories without topic shift in the low WM load condition elicited a sustained negative deflection relative to the other conditions, increasing as the

sentence unfolded. This slow shift was localized frontal and on the left side of the scalp, and has been associated with increased memory demands (a.o., Kutas & Federmeier, 2007; Bosch et al., 2001) and referential ambiguity (Nieuwland & van Berkum, 2006; van Berkum, 2008), as discussed in the Introduction.

The presence of this slow wave in the stories *without* topic shift might indicate that the repetition of a name causes more processing load because referring to a salient referent with a repeated name results in a less coherent discourse than using a pronoun instead (a.o., Gordon et al., 1993; Swaab et al., 2004). Interestingly, a series of studies (Swaab et al., 2004; Ledoux, Gordon, Camblin, & Swaab, 2007; Camblin, Gordon, & Swaab, 2007) that have compared ERPs on repeated names with ERPs on new names do not report a similar slow shift for repeated names, but only report a negative deflection up to 700 ms after the onset of the proper name. These studies used single sentences to test the effects of repetition of referring expressions, whereas in our materials the referents are mentioned in different sentences.

Regardless of what underlying language processing strategy may cause the slow wave shift, the absence of this shift in the high WM load condition suggests that participants change their on-line language processing when WM capacity is insufficient due to higher WM load. Note that we did not find a similar interaction in the N400 effects on reading the first proper name, which only shows two main effects of topic shift and WM load. Because the distribution of the N400 effects on reading the first proper name are clearly different from the distribution of the effects on reading the second proper name, we hypothesize that WM load may affect sentence processing in different ways: 1) higher WM load causes more difficulties in using information from the preceding context, as suggested by lower N400 amplitudes as discussed before, and 2) people change their language processing strategies if they have insufficient WM load available. Both effects of WM load are likely to affect story comprehension.

4.5 Story comprehension

On the basis of the effects in Sentence 2 and Sentence 4, we propose that the saliency of referents in the discourse already differs as a result of the grammatical position in which they are introduced, and this difference in saliency causes an early commitment of participants for one of the referents as being the discourse topic. WM load may influence this early difference in saliency, by attenuating the effect of the preceding discourse. Furthermore, our results indicate that higher WM load may change the strategy of sentence processing to avoid additional WM load. These effects of WM load will in turn affect story comprehension. Thus, we propose that the ERP effects found in Sentence 2 are predictive for off-line story comprehension. In contrast, the differences in saliency that arise already at the topic shift in Sentence 2, disambiguate the on-line pronoun resolution. Together, these results suggest that *off-line* pronoun resolution, necessary for story comprehension, may be guided by additional factors than *on-line* pronoun resolution.

Note that different factors, such as grammatical role, verb biases, and world-knowledge, influence the referent's discourse saliency. As a result, some discourses do not unambiguously establish a discourse topic. A pronoun following such a discourse structure will reflect the

ambiguity of the discourse structure and be truly ambiguous for the reader or listener.

4.6 Conclusion

We examined whether WM load and discourse structure interact in the on-line resolution of pronouns by measuring EEG during story comprehension in a dual-task experiment. The secondary task increased the WM load of participants in the high WM load condition while they read short stories. In contrast to other reading time studies, but in line with our previous reading time study using the same stories, we measured only small differences due to WM load and the discourse structure on reading the pronoun. However, earlier in the story, when a topic shift was elicited by introducing the second referent in subject position, we found significant effects of discourse structure and WM load, indicating that the discourse saliency of the referents resolved the potential ambiguity of the discourse. In addition, our data suggests that with higher WM load, participants use a different sentence processing mechanism that requires less WM capacity but may result in interpretational differences. On the basis of our results we conclude that a potential ambiguity is not necessary resolved at the onset of the pronoun, but might already be disambiguated at the introduction of the referents.

Notes

[1] All participants signed a form to document informed consent. In addition, undergraduate students younger than 18 received permission from parents or caregivers to participate in the experiment.

[2] We decided to include the left-handed participants in the data because of the within-subjects design of the study. We controlled for handedness by removing the time trends per subject from the data to reduce the noise from participants, and used the residual data for statistical analyses (as described in Section 3). After removing the individual time trends, no effect of handedness was found in the residuals.

Appendices

1 Statistical analyses

Package mgcv 1.7-19 (Wood, 2006) of the statistical software package R was used for analysis.

1.1 Procedure for removing variability due to participants and items

To avoid any spurious effects due to variability of participants and items, we used the following procedure of removing different sources of variation before analyzing the residual data for effects of Topic shift and WM load.

1. GAM model for removing main trend over time:
   Cz ~ s(Time, k=40)
2. GAM model for removing main trend over time for each participant:
   res0 ~ s(Time, by=Subject, k=40)
3. Linear mixed effect model for removing random intercept for each item:
   res1 ~ 1 + (1|Item)

1.2 Investigating the effects of Topic shift and WM load

Table A1: Best-fitting model for Cz over time window 1 (Sentence 2, word 1-2) on the residual data after removing the main trend of Time and variability due to subject and items.

Formula

```
res3 ~ s(Time, by = Shift) + s(Time, by = WM) + te(Time, Trial)
```

Predictor		Estimate	SE	t value	p-value
Intercept	***	-0.23007	0.05528	-4.162	<.001

Smooth term		edf	Ref.df	F	p-value
s(Time):Shift	***	7.571	8.673	22.234	<.001
s(Time):WM	***	6.808	7.941	22.701	<.001
te(Time,Trial)	***	14.554	18.132	7.827	<.001

Model characteristics

```
R-sq.(adj) = 0.00244, Deviance explained = 0.26%
REML score = 7.4211e+05, Scale est. = 179.72, n = 184842
```

Table A2: Best-fitting model for Cz over time window 2 (Sentence 2, word 3-4) on the residual data after removing the main trend of Time and variability due to subject and items.

Formula

```
res3 ~ s(Time, by = Shift) + s(Time, by = WM) + s(Time, by = Interaction) +
te(Time, Trial)
```

Predictor		Estimate	SE	t value	p-value
Intercept	***	-1.24779	0.08557	-14.58	<.001

Smooth term		edf	Ref.df	F	p-value
s(Time):Shift	***	6.082	7.205	19.069	.<001
s(Time):WM	***	3.062	3.568	68.502	<.001
s(Time):Interaction	***	2.094	2.168	37.911	<.001
te(Time,Trial)	**	13.788	17.370	2.345	0.00125

Model characteristics

```
R-sq.(adj) = 0.00195, Deviance explained = 0.208%
REML score = 7.9451e+05, Scale est. = 316.88, n = 184842
```

Table A3: Best-fitting model for Cz over time window 3 (Sentence 4, word 1-2) on the residual data after removing the main trend of Time and variability due to subject and items.

Formula

```
res3 ~ s(Time, by = Shift) + s(Time, by = WM) + s(Time, by = Interaction) +
te(Time, Trial)
```

Predictor		Estimate	SE	t value	p-value
Intercept	**	-0.19536	0.06409	-3.048	0.0023

Smooth term		edf	Ref.df	F	p-value
s(Time):Shift	**	4.752	5.627	3.821	0.0012
s(Time):WM	***	5.658	6.681	8.101	<.001
s(Time):Interaction	***	5.542	6.524	5.300	<.001
te(Time,Trial)	***	15.157	18.051	10.484	<.001

Model characteristics

```
R-sq.(adj) = 0.00179, Deviance explained = 0.196%
REML score = 7.2753e+05, Scale est. = 171.06, n = 182331
```

2 Distribution of effects

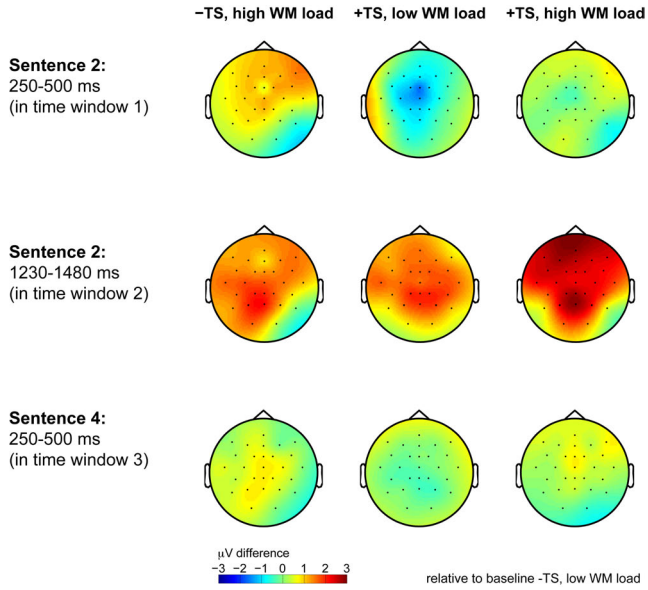


Figure A1: Estimated scalp distributions in N400 range for each time window. Averages are based on the predicted effects of the best fitting model for each electrode.

chapter 6

Event-related pupil dilation tracks integration of contextual and grammatical information

Abstract

This study measured pupil dilation to investigate whether and when during sentence processing listeners use contextual information to resolve object pronouns (*him*, *her*). The interpretation of object pronouns is assumed to be constrained by grammatical principles. However, in this study we found that the referents' prominence in the linguistic and visual context also affects pupil dilation. To manipulate the discourse prominence of the referents, we changed the order of introduction of the two referents. When the agent of the target sentence with the object pronoun was introduced first in the linguistic context, the visual context affected pronoun resolution. On the other hand, it did not when the agent was introduced second. Very early after the onset of the pronoun, pupil dilation showed reliable differences between these conditions, indicating that linguistic and visual context compete with grammatical principles during pronoun resolution.

1 Introduction

Pronouns such as *he* and *him* differ from proper names like *John* in that they can refer to various referents in the linguistic or non-linguistic discourse. To interpret pronouns, listeners often base their choice of referent on features of the preceding linguistic discourse that they have stored in memory. These features concern, among other things, the grammatical roles of the referents in the linguistic discourse (a referent that is the subject of the previous utterance is preferred over a referent that is an object of the previous utterance), order of mention (the referent that is mentioned first is preferred over a referent that is mentioned later in the linguistic discourse), and recency (referents that were referred to in a recent utterance are preferred over referents that were referred to several utterances back) (see Arnold, 1998, for an overview). The effects of these determinants of discourse prominence on pronoun resolution have been established on the basis of the interpretation of subject pronouns (*he*, *she*).

In contrast, linguistic discourse is often assumed to play only a minor role in the interpretation of object pronouns (*him*, *her*). In languages such as English and Dutch, object pronouns cannot have the same reference as the subject of the clause, as is illustrated by sentence pair 1:

1. Here you see a penguin and a sheep. The penguin is hitting him with a pan.

In this sentence pair, the object pronoun *him* cannot refer to the same individual as the subject of the same clause, the penguin, and hence is preferably interpreted by adult listeners as referring to the other referent in the linguistic discourse, the sheep. This restriction on the interpretation of object pronouns is known in linguistic theory as Principle B of Binding Theory (Chomsky, 1981). Because the interpretation of object pronouns is guided by Principle B of the grammar, it is often assumed that the linguistic discourse only plays a minor role, if any, in the interpretation of object pronouns. However, several recent studies suggest that the linguistic discourse does influence children's offline interpretation of object pronouns as well as adults' online interpretation of object pronouns (Clackson, Felser, & Clahsen, 2011; Spenader et al., 2009).

This raises the question whether, how and when information from the linguistic discourse is taken into account in the interpretation of object pronouns. Do listeners initially only apply grammatical principles to guide their search for possible referents? Or do listeners also use information from discourse in object pronoun resolution, as they do in subject pronoun resolution? And if so, at which point in time is the information from grammar and discourse combined? The reported study investigates when listeners use the linguistic discourse in their resolution of object pronouns.

1.1 Previous research on the on-line resolution of object pronouns

How do listeners combine different sources of information during sentence processing to determine the interpretation of an object pronoun? To explain the time course of pronoun resolution, two conflicting types of accounts have been proposed. The *initial-filter account* (a.o., Clifton, Kennison, & Albrecht, 1997; Nicol & Swinney, 1989) assumes that the set of poten-

tial antecedents for the pronoun is restricted by the grammar. Other sources of information, such as discourse context, only affect the choice of antecedent at a later stage of processing, when the grammar has already filtered out antecedents that violate Principle B. In contrast, the *competing-constraints account* (a.o., Badecker & Straub, 2002; Kennison, 2003) assumes that grammar competes with other sources of information during pronoun resolution. Hence, discourse context may affect the choice of antecedent for a pronoun already from the start.

Experiments that addressed whether the grammar acts as an initial filter on the set of potential antecedents for an object pronoun yielded mixed results. For example, Nicol and Swinney (1989) used cross-modal priming to investigate the reactivation of referents in sentences like 2:

2. The boxer told the skier that the doctor for the team would blame him • for the recent injury. (• indicates the probe point)

They found significant priming effects after the pronoun *him* only for the *boxer* and the *skier*, but not for the local subject *the doctor*, suggesting that only grammatically appropriate antecedents are reactivated during pronoun resolution. In a series of self-paced reading studies, however, Badecker and Straub (2002) found opposite effects. In their studies, they focused on the referent that is ruled out as the antecedent of the pronoun by Principle B of the grammar because it is the subject of the same clause. Sentences in which this local subject matches the object pronoun *him* in gender (e.g., if the local subject would be *John*) are compared with minimally differing sentences in which the local subject does not match the object pronoun in gender (e.g., if the local subject would be *Jane*). Badecker and Straub report that reading times are increased on the words following the pronoun when the local subject matches the pronoun in gender (*John*) compared to when the local subject does not match the pronoun in gender (*Jane*). This difference in reading times suggests that a local matching subject is at least initially considered as a potential antecedent of the pronoun. If the local subject were already filtered out by the grammar as a potential antecedent of the pronoun, as stated by the initial-filter account, it should not matter whether the gender of the local subject matches the pronoun or not.

A recent study by S. Lewis, Chow, and Phillips (2012) using the same methodology, however, was unable to replicate Badecker and Straub's findings. This contrasts with a visual-world eye-tracking study by Clackson et al. (2011), whose results seem to be in line with Badecker and Straub (2002). Clackson et al. (2011) investigated the time course of listeners' interpretation of object pronouns by presenting participants with two-sentence stories such as Example 3.

3a. Peter was waiting outside the corner shop. He watched as Mr. Jones bought a huge box of popcorn for him over the counter.

3b. Susan was waiting outside the corner shop. She watched as Mr. Jones bought a huge box of popcorn for her over the counter.

Participants showed significantly fewer looks to Peter in the presence of a local subject of the same gender (*Mr. Jones* in sentence 3a) compared to looks to Susan in the presence of

a local subject of a different gender (*Mr. Jones* in sentence 3b). This result confirms Badecker and Straub's finding that local subjects are not immediately filtered out by the grammar as potential antecedents of an object pronoun. All in all, therefore, previous work has yielded contradictory results regarding the role of grammatical information.

In this study, we investigate how and when the linguistic discourse and grammar interact during the resolution of object pronouns. We do not manipulate the gender properties of potential antecedents, as previous studies did, but rather we manipulate their prominence in the discourse. Under the initial-filter account as well as the competing-constraints account, determinants of discourse prominence such as the order of mention could influence the interpretation of object pronouns too, just like they influence the interpretation of subject pronouns. However, the accounts make different predictions about the effect of discourse prominence: Under the competing constraints account but not the initial filter account, discourse prominence also affects the availability of the local subject as a potential antecedent of the pronoun. We thus aim to investigate whether and when discourse factors such as order of mention influence the availability of the local subject as the antecedent of an object pronoun in simple transitive sentences. We investigate these issues using the method of pupil dilation, which is a robust measure to detect differences in processing load during on-line language processing (a.o., Beatty & Lucero-Wagoner, 2000; Just & Carpenter, 1993; Engelhardt, Ferreira, & Patsenko, 2010).

1.2 Effects of discourse prominence

If the linguistic discourse influences object pronoun resolution even at the earliest moments of processing the pronoun, we may find signatures of the same discourse factors that influence subject pronoun interpretation, such as order of mention (e.g., Gernsbacher & Hargreaves, 1988; Gordon et al., 1993; Kaiser & Trueswell, 2008). Firstly mentioned referents are perceived as more prominent and hence are more likely antecedents of a subject pronoun than referents that are mentioned later. In the reported study, we manipulate the discourse prominence of the local subject by changing the order of introduction in sentences such as Example 1. Recall that Principle B of the grammar prohibits the local subject *the penguin* as the antecedent of the pronoun in a sentence such as *The penguin is hitting him with a pan.* Reversing the order of introduction of the two referents (i.e., *Here you see a sheep and a penguin* rather than *Here you see a penguin and a sheep*) and introducing the penguin secondly rather than firstly makes *the penguin* less prominent. If this first mention bias not only influences the interpretation of subject pronouns but also of object pronouns, the reversed order of introduction is expected to result in less processing difficulty. That is, decreasing the discourse prominence of the local subject is expected to have a similar effect as replacing a local subject that matches the pronoun in gender by a referent that does not match the pronoun.

To investigate whether and when discourse prominence affects the processing of object pronouns, we performed a visual-world eye-tracking study in which we measured the participants' pupil size during the task.

1.3 Pupil dilation

The size of the pupil is mainly determined by light and accommodation reflexes, but in conditions of constant lighting, small fluctuations (usually less than 0.5 mm) in pupil size can be measured that have been found to reflect cognitive processing (for reviews, see, a.o., Beatty, 1982; Beatty & Lucero-Wagoner, 2000; Janisse, 1977). These fluctuations in pupil size have been found to be a reliable measure of processing load in tasks that investigate memory, attention, reasoning, decision-making and memory. Pupil dilation is a slow measure that peaks around 1000 ms after the stimulus that triggered the dilation. However, the peak may vary between tasks: For example, Hoeks and Levelt (1993) estimated the mean peak latency in a simple reaction task on 930 ms after the stimulus, whereas Just and Carpenter (1993) report peak latencies around 1.3 ms.

Pupil dilation is not used very often to study language comprehension, although pupil dilation is sensitive to effects of processing load for various linguistic phenomena (Engelhardt et al., 2010; Hyönä, Tommola, & Alaja, 1995; Just & Carpenter, 1993; Scheepers & Crocker, 2004; Zellin, Pannekamp, Toepel, & van der Meer, 2011). For example, Just and Carpenter (1993) compared subject and object relative clauses and different types of filler-gap sentences. Their results show more pupil dilation for more complex sentences.

More recently, Engelhardt et al. (2010) reported that prosody affected pupil dilation during the processing of syntactic ambiguities if the presented visual context was inconsistent with the correct interpretation of the sentence. Zellin et al. (2011) investigated how discourse information (new versus contrastive information) and prosody influenced processing load. Different effects were found for peak dilation, the maximal dilation of the pupil, and peak latency, the amount of time between the start of the measurement interval and the peak dilation.

These studies show that pupil dilation is sensitive to subtle linguistic differences, associated with differences in processing load. Therefore, we measured pupil dilation during sentence processing to investigate whether the discourse prominence of the local subject referent influences processing load.

2 Experimental materials and methods

The current study aimed to investigate whether and when referents' discourse prominence affects processing load during the resolution of object pronouns. The presence of an immediate effect of the local subject's prominence on pupil dilation would suggest that the grammar does not filter out the local subject as a potential antecedent, but that the linguistic discourse competes with the grammar during object pronoun processing. We manipulated the discourse prominence of the local subject referent by changing the order of introduction of the two referents.

2.1 Materials and task

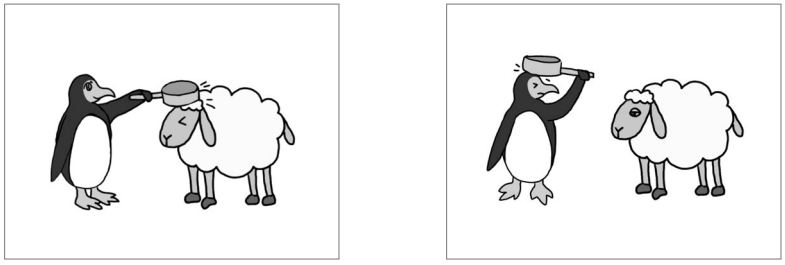
To study object pronoun resolution, we used a picture verification task in a visual-world paradigm. Participants were asked to judge whether a two-sentence story that was auditorily presented correctly described the picture presented on the screen. Sentences were pre-recorded and presented starting 500 ms after the picture appeared on the screen.

In the first sentence, two referents were introduced (see Example 4). The order of introduction was manipulated to change the discourse prominence of the referents. With an Agent-Patient (AP) introduction the agent becomes more prominent in the discourse because it is introduced first in the discourse (first mention bias, a.o., Gernsbacher & Hargreaves, 1988). In contrast, with a Patient-Agent (PA) introduction the prominence of the patient increases because the patient is introduced first in the discourse.

4. Example test sentences (in Dutch, with their English translations given below):

Introduction Order *Agent-Patient* (AP):	Hier zie je een pinguïn en een schaap. 'Here you see a penguin and a sheep'
Introduction Order *Patient-Agent* (PA):	Hier zie je een schaap en een pinguïn. 'Here you see a sheep and a penguin'
Test sentence:	De pinguïn slaat hem met een pan. 'The penguin is hitting him with a pan'

Figure 1 shows two pictures for which the test sentence in 4 is a correct or an incorrect description. Instead of choosing between these two pictures (i.e., a picture-selection task), participants were presented with only one picture and were asked to judge whether the presented sentence was a correct description of the picture or not. If the picture on the screen showed an other-oriented action (Figure 1a), the description was correct (*congruent item*). If the picture showed a self-oriented action (Figure 1b), the description was incorrect (*incongruent item*). In addition, filler sentences without a pronoun were included in the experiment to provide incorrect descriptions for the pictures with an other-oriented action and correct



a. Picture showing other-oriented action b. Picture showing self-oriented action

Figure 1: A pictures that is congruent (left) and a picture that is incongruent (right) with the sentence 'The penguin is hitting him with a pan'.

descriptions for pictures with a self-oriented action.

Each trial started with a fixation cross. After 750 ms the picture appeared on the screen. In half of the trials the picture was mirrored (randomly determined before each trial). The pre-recorded sentences started 500 ms later. The picture remained visible on the screen another 2500 ms after the sentences were finished, so that the pupil size could decrease to its normal size. The picture on the screen was then replaced by the two possible answers ('right', 'wrong') presented on the left and right side of the screen. To prevent participants from preparing their motor response, the location of the answer on the screen was randomized per trial. This delayed and randomized presentation of the answer options allows us to disentangle the pupil dilation that is triggered by sentence processing from the pupil dilation that is triggered by preparing the response on the judgment task. After an answer was given using one of the two front buttons of a game pad, a blank screen was presented for 2500 ms before the next trial started.

2.2 Design and procedure

The 2x2 design (Introduction Order x Congruency) was tested within subjects. Four variants were created of each of the 32 test items that were distributed over 4 lists. In addition, 32 filler items were created analogously to the experimental items, but without a pronoun. Instead, half of the fillers contained a reflexive object and the other half of the fillers were intransitive sentences that did not contain an object. The filler items were balanced for type of Introduction Order and Congruency. All items ended with a prepositional phrase.

The pupil area was monitored continuously during the picture verification task with an EyeLink 1000 (SR research) at 250 Hz (settings: monocular tracking based on pupil and corneal reflection, noise reduction with filter level 2). Stimulus presentation was programmed using E-Prime 2.0 software. The eye-tracker was calibrated at the start of the experiment and in the two breaks, at 1/3 and 2/3 of the experiment. The computer screen was set at a resolution of 1,024 x 786. The background of the screen was white during the trials and calibration. A chin and forehead rest was used to keep the distance between the participant's head and the screen constant. Luminance of the room was normal and kept constant during the experiment.

2.2.1 Participants

Participants were 18 native-Dutch speaking students, who were paid 10 euros or rewarded with course credits. One participant was excluded because of blinks and track loss in more than half of the trials. The data of 17 participants (mean age 19.4, range 18-21; 4 men) is analyzed.

2.2.2 Pupil dilation recording and processing

Eye blinks were corrected for by linear interpolation. Trials with more than 25% of the data being replaced by interpolation were excluded from analyses (n=14 of 497). The data was filtered with a low pass Butterworth filter with an upper band of 25 Hz using the package 'signal'

of the statistical software package R (www.r-project.org). Trials were temporally aligned to the onset of the pronoun, and down sampled to 50 Hz. The average pupil size in the interval from 0 ms to 250 ms after the picture onset (i.e., 500 to 250 ms before the sentence onset) was used as a baseline. Data were normalized by calculating the proportion of increase of the pupil size compared to the baseline for each data point.

3 Results

3.1 Off-line performance

Participants gave very few wrong answers on the task. Answer accuracy was more than 93%. We excluded incorrectly answered items (n=16 of 483) from all subsequent analyses.

3.2 Pupil dilation

Figure 2 shows the average pupil dilation per condition (Introduction Order x Congruency), aligned on the onset of the pronoun, from the start of the trial to approximately 3000 ms after the pre-recorded sentences are finished.

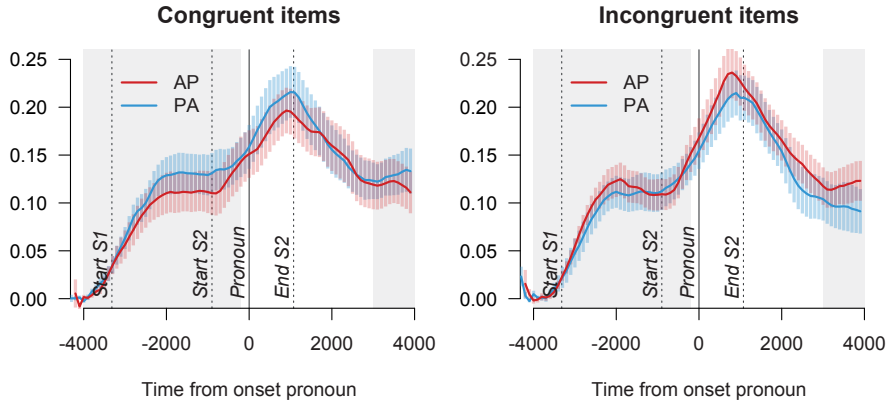


Figure 2: Mean pupil dilation (±1SE) from the start of the trial to 4000 ms after the onset of the pronoun (baseline 0 to 250 ms after the start of the trial) for the congruent trials (with picture showing other-oriented action) and incongruent trials (with picture showing self-oriented action), averaged over time bins of 100 ms for presentation purposes. The red lines indicate an agent-patient (AP) introduction sentence, and blue lines a patient-agent (PA) introduction sentence. Vertical lines show the average onset of sentence 1 and sentence 2, and the average offset of sentence 2.

In the introduction sentence, the pupil dilation increases until the introduction of the second referent. From the start of the test sentence, the dilation starts to increase again up to the end of the sentence, after which the dilation decreases. To test for effects of referent

introduction on pronoun resolution, the pupil dilation is analyzed from 200 ms before the onset of the pronoun to 3000 ms after the pronoun (the white areas in Figure 2), aligned on the onset of the pronoun and with the same baseline of 0 to 250 ms after the start of the trial.

Previous studies report differences in (mean) peak dilation, peak latency (a.o., Beatty & Lucero-Wagoner, 2000; Just & Carpenter, 1993) and sometimes dilation slope (e.g., Engelhardt et al., 2010). In this study, we use generative additive models (GAMs, Wood, 2006; see Tremblay & Baayen, 2010 for an application of GAMs to EEG data from a immediate free recall experiment, and Wieling, Nerbonne, & Baayen, 2011, for an application of GAMs to investigate linguistic variation) to analyze pupil dilation because this approach can capture differences in peak dilation and peak latency within a single analysis.

3.3 The effect of order of introduction and congruency on pupil dilation

GAMs are generalized linear models in which linear predictors can be specified by non-linear functions of predictor variables (i.e., *smooth functions*). These functions are estimated using a penalized regression method called penalized iteratively re-weighted least squares, and the parameters for each smooth function such as the degree of smoothness are estimated using cross-validation (see Wood, 2006, chapters 3 and 4). These estimation processes determining the smooth functions and parameters are designed to avoid overgeneralization and overfitting of the data.

We use GAMs to estimate the effects of Congruency and Introduction on the time-course of the un-averaged pupil dilation. To avoid any spurious effects due to variability of participants and items, we used a conservative procedure of removing different sources of variation and analyzing the residual data for effects of Congruency and Introduction. First, we removed the main trends of pupil dilation over time, followed by the time trends of individual subjects. Subsequently, we removed the time trends of individual items, and also included a predictor to account for the variability in the time trends due to order of trials in the experiment, and a linear predictor for the effect image direction, which pertains to the positions of agent and patient in the picture. To investigate whether the residual data is influenced by Congruency and Introduction Order over time, we included smooth functions to account for the time trends for each condition. We compared different models using a stepwise variable deletion procedure, starting with the complete interaction model, which was the best fitting model (Appendix 1, Table 4). The best fitting model includes an intercept for *Condition* (factor with four levels representing the Introduction Order x Congruency conditions; $F(1.26,74704)=5.27$; $p=.015$) and the effect of Condition over time ($F(12.43, 74704)=11.93$; $p<.001$), providing evidence that Introduction Order and Congruency influence the measured pupillary dilation. Figure 3 shows how these factors influence the pupil dilation.

The panel in the top row of Figure 3 shows the estimated effect of Time from the onset of the pronoun. The pupil dilation starts with an intercept of 0.165 and peaks around 1000 ms after the pronoun onset (indicated by the dotted vertical line), which is close to the peak latency of 930 ms reported by Hoeks and Levelt (1993).

The second row of Figure 3 shows in the four rightmost panels the estimated effects for

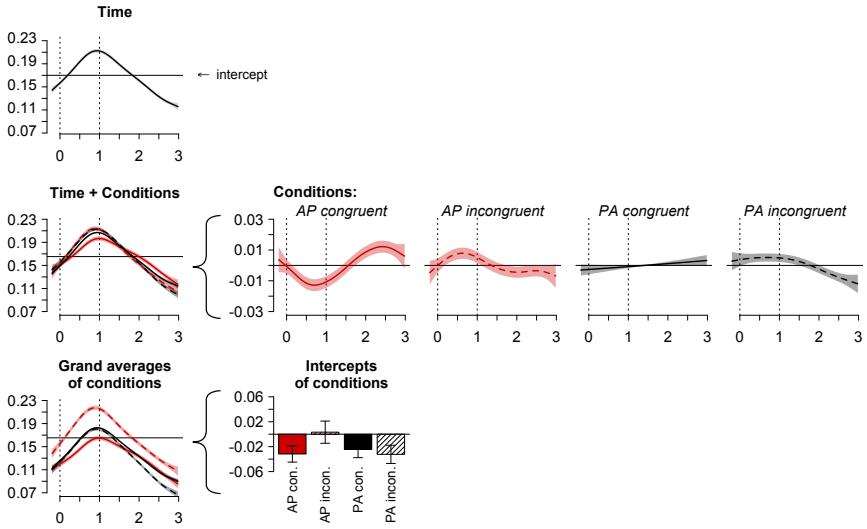


Figure 3: Estimated effects (± 1 SE) of the predictors on pupil dilation over time from the onset of the pronoun. *First row:* the main effect of Time peaking around 1000 ms after the onset of the pronoun as indicated by the dashed line. *Second row:* the four rightmost panels show the partial effects of Time x Condition; the leftmost panel shows the effects of condition added to the main effect of time. *Third row:* The right panel shows the difference in intercepts of each condition, indicating the differences in pupil dilation at the start of the pronoun. The left panel shows the effects of Condition over time with intercepts added.

each condition (Introduction Order x Congruency) over Time, after removing the main time trend of pupil dilation. In the AP-congruent condition, the pupil dilation peak is reduced, indicated by the initially negative curve. However, the pupil dilation peak is increased in the incongruent conditions AP-incongruent and PA-incongruent, which show a similar initially positive curve. The PA-congruent condition does not show a significant difference with the main time trend, indicated by the flat line around zero. Note also that the end of the curves for the incongruent items (AP-incongruent and PA-incongruent) is negative, suggesting a faster decrease of pupil dilation, but positive for the AP-congruent condition, suggesting a slower decrease of pupil dilation. The leftmost panel shows the effects of condition added to the main time trend. The peak dilation is considerable lower in the AP-congruent condition than in all other conditions, and it takes longer in the AP-congruent condition for the pupil dilation to reduce to the baseline. The peak dilation in the PA-congruent condition is somewhat lower than in the incongruent items. The differences between conditions arise early in the time course: within 500 ms after the pronoun onset the dilation slopes separate. There was no difference over time between the AP-incongruent condition and the PA-incongruent condition from the onset of the pronoun.

The bottom row completes the pattern of differences between conditions over time. The

right panel of the bottom row shows the estimated intercepts per condition, reflecting the difference from the mean in pupil size at the onset of the pronoun. Treatment coding was used, which contrasts a baseline condition (AP-congruent; β=-.032, SE=.013, t=-2.43, p=.015) with other conditions. Only AP-incongruent has a higher intercept than AP-congruent (β=.034, SE=.018, t=1.97, p=.049).

The rightmost panel of the bottom row shows the effects of each condition over time added to the condition intercepts. If the local subject is very prominent in the linguistic context (i.e., in the AP conditions), then the visual context makes a difference in peak dilation: if the picture shows a self-oriented action, participants show significantly increased processing load compared to the other conditions, but if the picture shows an other-oriented action, participants show reduced processing load. The visual context does not influence processing load much if the local subject referent is introduced second (i.e., in the PA conditions).

To summarize, not only the order of referent introduction (Introduction Order), but also the visual context (Congruency) influences pupil dilation during on-line pronoun resolution.

4 Discussion

To investigate whether and at which point during sentence processing listeners use discourse prominence in their resolution of object pronouns, we carried out a visual-world picture verification task with Dutch adults, while measuring their pupil size. We manipulated discourse prominence through the order of referent introduction in the sentence preceding the object pronoun. Participants showed an interaction between the type of picture and the order of referent introduction. According to the grammar of Dutch, a non-reflexive object pronoun cannot refer to the same referent as the local subject. When this 'ungrammatical' antecedent was introduced before a competing 'grammatical' antecedent and hence was more prominent, participants experienced more processing load in the incongruent items than in the congruent items. If the grammatical antecedent was introduced first, the congruency effect was much smaller.

4.1 Theoretical implications

According to the initial-filter account, the grammar acts as a filter removing ungrammatical antecedents as potential referents of the object pronoun. This implies that only referents that are allowed by the grammar as the antecedent of the pronoun are considered during on-line pronoun resolution. The local subject is not allowed by the grammar and hence its discourse prominence should not influence the resolution of the object pronoun. The competing-constraints account, on the other hand, assumes that grammar and discourse compete during pronoun resolution. This implies that not only properties of grammatical referents but also of ungrammatical referents may influence on-line pronoun resolution.

The interaction between grammar, linguistic discourse and visual context found in this study cannot be explained in terms of an initial-filter account. According to the initial-filter account, the subject of the test sentence is not a grammatically appropriate antecedent of

the object pronoun and hence is not even considered. However, when this ungrammatical antecedent is prominent in the linguistic discourse because it is introduced first and, in addition, the visual context is congruent with the sentence (as in the AP-congruent condition), lower pupil dilation is measured than when the ungrammatical antecedent is less prominent (as in the PA-congruent condition). This is interpreted as lower processing load in the AP-congruent condition than in the PA-congruent condition. When the visual context is incongruent with the sentence, the order of introduction has an opposite effect on pupil dilation. Thus, these results suggest that the grammar, linguistic discourse and visual context interact during the on-line resolution of object pronouns. Moreover, the linguistic discourse and visual context influence pupil dilation very early in the pronoun resolution process because the conditions start to differ in pupil dilation slope quickly (around 250 ms) after the onset of the pronoun.

4.2 The influence of contextual information

Engelhardt et al. (2010) report a similar interaction between visual context and prosody in the processing of garden path sentences. As discussed in the introduction, they found that prosodic information reduced processing load if it was congruent with the correct reading of a temporarily syntactically ambiguous sentence, but only if the visual context was incongruent with this correct interpretation. If the picture matched the correct, but initially less preferred, reading of the sentence, prosody did not affect processing load because the visual context already may have disambiguated the sentence. In our experiment, the visual context could not disambiguate the sentence before the onset of the pronoun, because the sentence could have continued with a reflexive (*himself*) rather than a pronoun, resulting in a match between the sentence and the picture. Therefore, the effects of visual context are somewhat different in our study than in Engelhardt et al.'s.

Although the visual context could not disambiguate the sentence in our study, participants may have relied on the accompanying picture more in the AP conditions than in the PA conditions. In the AP-congruent condition, pronoun processing is facilitated by the visual context, indicated by lower pupil dilation. In contrast, in the AP-incongruent condition pronoun processing is hindered by the conflicting information from the visual context. Generally, speakers tend to mention the agent before the patient or place the agent in a more prominent grammatical position (e.g., the subject position) than the patient when describing pictures (a.o., Griffin & Bock, 2000; Gleitman, January, Nappa, & Trueswell, 2007; Myachykov, Thompson, Scheepers, & Garrod, 2011). As a consequence, a discourse starting with an AP-introduction is more natural than a discourse starting with a PA-introduction, given the picture on the screen. Following this reasoning, the PA-introduction may warn the participants not to rely on the interpretation presented on the screen because the PA-introduction suggests that the patient is more important than the visual context shows, which conflicts with the information of the visual context (in both the congruent and incongruent conditions). This could explain the lack of influence of visual context on the PA conditions.

This explanation, which proposes that the visual context interacts with linguistic infor-

mation during sentence processing, is in line with previous eye-tracking studies (a.o., Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Knoeferle, Crocker, Scheepers, & Pickering, 2005) showing that the visual context resolves or reduces temporarily syntactic ambiguities during on-line sentence processing. These effects of visual context provide support for the competing constraints account, which assumes that non-linguistic information may influence on-line pronoun resolution already early on.

4.3 Conclusion

Using pupil dilation as a measure of processing load, we found effects of linguistic discourse (the order of referent introduction) and visual context (whether the sentence and the picture match or do not match) during the resolution of object pronouns in a sentence-picture verification task. Together, these results provide strong evidence against the view that the grammar acts as an initial filter on the selection of possible antecedents for object pronouns. Our study thus supports the view that grammar, linguistic discourse and visual context compete in determining the antecedent of an object pronoun.

Appendices

1 Statistical analyses

Package mgcv 1.7-18 (Wood, 2006) of the statistical software package R was used for analysis.

Table A1: Model 1 for removing the main trend over time from the onset of the pronoun.

Formula					
`Pupil ~ s(Time)`					
Predictor		**Estimate**	**SE**	**t value**	**p-value**
`Intercept`	***	0.165	0.000625	264.3	<.001
Smooth term		**edf**	**Ref.df**	**F**	**p-value**
`s(Time)`	***	7.248	8.265	271.6	<.001
Model characteristics					
`R-sq.(adj) = 0.0292, deviance explained = 2.93%`					
`REML score = -25950, scale estimation = 0.029215, n=74720`					

Table A2: Model 2 for removing the main trend over time per subject from residuals of model 1.

Formula

```
res1 ~ s(Time, by=Subject)
```

Predictor		Estimate	SE	t value	p-value
Intercept	***	0.902	0.037	24.11	<.001
Smooth term		**edf**	**Ref.df**	**F**	**p-value**
s(Time):Subject1	***	5.722	6.733	43.522	<.001
s(Time):Subject10	***	5.131	6.075	25.504	<.001
s(Time):Subject11	***	7.311	8.235	181.826	<.001
s(Time):Subject12	***	7.025	8.002	38.138	<.001
s(Time):Subject13	***	4.684	5.553	46.846	<.001
s(Time):Subject14	***	5.288	6.253	18.312	<.001
s(Time):Subject15	***	7.241	8.18	146.61	<.001
s(Time):Subject16	***	7.738	8.539	382.204	<.001
s(Time):Subject17		1.019	1.032	0.211	0.654
s(Time):Subject2	***	3.647	4.356	5.111	0.000287
s(Time):Subject3	***	5.721	6.733	17.72	<.001
s(Time):Subject4	***	7.426	8.322	500.588	<.001
s(Time):Subject15	***	6.302	7.334	153.513	<.001
s(Time):Subject16	***	7.213	8.157	226.5	<.001
s(Time):Subject17	***	5.053	5.985	17.981	<.001
s(Time):Subject18	***	6.874	7.871	210.374	<.001
s(Time):Subject19	***	3.512	4.203	27.966	<.001

Model characteristics

```
R-sq.(adj) = 0.237, deviance explained = 23.8%
REML score = -35856, scale estimation = 0.022286, n=74720
```

Table A3: Model 3 for removing the main trend over time per item from residuals of model 2, and the effects of trial position and whether the picture was mirrored.

Formula

```
res2 ~ ImgDir + s(pTime, by = Item) + te(pTime, trial)
```

Predictor		Estimate	SE	t value	p-value
Intercept	***	0.607	0.0795	7.64	<.001
ImgDir	***	-0.0136	0.000529	-25.75	<.001

Smooth term		edf	Ref.df	F	p-value
s(Time):Item	***	4.143	4.914	55.43	<.001
s(Time):Item10	***	4.722	5.598	23.024	<.001
s(Time):Item11	***	5.315	6.285	61.384	<.001
s(Time):Item12	***	4.682	5.552	34.304	<.001
s(Time):Item13	***	5	5.924	5.834	<.001
s(Time):Item14	***	6.272	7.304	206.779	<.001
s(Time):Item15	***	3.99	4.735	15.209	<.001
s(Time):Item16	***	4.804	5.695	17.547	<.001
s(Time):Item17	***	4.425	5.246	21.409	<.001
s(Time):Item18	***	3.962	4.703	8.249	<.001
s(Time):Item19	***	3.83	4.551	24.287	<.001
s(Time):Item2	***	4.589	5.441	12.583	<.001
s(Time):Item20	***	6.099	7.131	17.636	<.001
s(Time):Item21	***	4.429	5.251	13.338	<.001
s(Time):Item22	***	4.261	5.053	21.735	<.001
s(Time):Item23	***	6.011	7.039	121.698	<.001
s(Time):Item24	***	5.075	6.012	15.705	<.001
s(Time):Item25	***	3.629	4.32	25.384	<.001
s(Time):Item26	***	4.644	5.506	7.072	<.001
s(Time):Item27	***	5.668	6.676	21.044	<.001
s(Time):Item28	**	1	1	9.894	0.00166
s(Time):Item29	***	5.703	6.713	22.474	<.001
s(Time):Item3	***	3.352	4	24.887	<.001
s(Time):Item30	***	4.858	5.758	28.093	<.001
s(Time):Item31	***	4.52	5.357	34.935	<.001
s(Time):Item32	***	4.502	5.338	29.346	<.001
s(Time):Item4	***	4.053	4.808	24.43	<.001
s(Time):Item5	***	4.376	5.189	9.626	<.001
s(Time):Item6	***	4.495	5.33	10.92	<.001
s(Time):Item7	***	2.826	3.371	23.205	<.001
s(Time):Item8	***	5.238	6.198	92.289	<.001
s(Time):Item9	***	4.061	4.817	15.201	<.001
te(Time,trial)	***	19.588	21.599	132.865	<.001

Model characteristics

```
R-sq.(adj) = 0.145, deviance explained = 14.7%
REML score = -41634, scale estimation = 0.01902, n=74720
```

Table A4: Best fitting model explaining the effects of Introduction Order and Congruency on pupil dilation from the onset of the pronoun (on the residuals of model 3).

Formula

```
res3 ~ Condition + s(pTime, by = Condition)
```

Predictor		Estimate	SE	t value	p-value
(Intercept)	*	-0.0317	0.0131	-2.430	0.0151
ConditionPA-congruent		0.00755	0.0135	0.561	0.5747
ConditionAP-incongruent	*	0.0350	0.0178	1.965	0.0494
ConditionPA-incongruent		-0.000601	0.0146	-0.041	0.9672

Smooth term		edf	Ref.df	F	p-value
s(Time):Cond.AP-congruent	***	4.596	5.637	14.079	<.001
s(Time):Cond.PA-congruent		1.045	1.089	0.315	0.594
s(Time):Cond.AP-incongruent	***	4.169	5.139	4.750	<.001
s(Time):Cond.PA-incongruent	***	2.618	3.261	9.795	<.001

Model characteristics

```
R-sq.(adj) = 0.0049, deviance explained = 0.51%
REML score = -42219, scale estimation = 0.018885, n=74720
```

chapter 7

Discussion

1 Introduction

This thesis investigated the relative contribution of linguistic constraints and cognitive factors in pronoun processing. I implemented two cognitive models in the cognitive architecture ACT-R to explain how children acquire, and adults process, pronouns in subject and object position. These cognitive models incorporated linguistic constraints restricting the production and interpretation of referring expressions: pronouns such as the subject form *he* and the object form *him*, reflexives such as *himself*, and definite full NPs such as the definite description *the pirate* and the proper name *Eric*. Based on computational simulations of the cognitive models and the results of various empirical and psychophysiological studies with Dutch-speaking children and adults, I argue that adult-like pronoun processing requires sufficient processing speed to take into account the perspective of the conversational partner. In addition, adult-like pronoun processing requires sufficient working memory (WM) capacity to build and update a representation of the linguistic discourse. In this discussion, I will first return to the main research questions posed in the introduction. Then, I will discuss the implications of this investigation for language acquisition, language processing and language disorders.

2 Effects of cognitive factors and linguistic constraints

Chapter 2 addressed children's acquisition of object pronouns in languages such as Dutch and English. In these languages, children up to 7 years old still make errors in their interpretation of object pronouns such as *him*, although at the same time they show adult-like production of object pronouns. I tested the hypothesis that children's interpretation follows from the constraints of the grammar, which allow object pronouns to corefer with the subject. Adult listeners take into account the speaker's perspective and thus block coreference between the object pronoun and the subject, but children are not capable of such perspective taking. On the basis of computational simulations of our cognitive model, I predicted that children are not yet able to complete the process of perspective taking because they lack sufficient processing speed. The results from our empirical study with Dutch-speaking 4- to 6-year-old children provide support for our cognitive model. When presenting children with pronoun sentences in slowed-down speech and normal speech, children's interpretation of object pronouns improved with slowed-down speech. On the other hand, their interpretation of reflexives (*himself*) deteriorated with slowed-down speech. This selective effect of slowed-down speech suggests that children who make errors with object pronouns do so because they fail to complete the necessary process of perspective taking at normal speech rate.

In **Chapter 3** another cognitive model was discussed. This model is in fact an extension of the cognitive model presented in Chapter 2. The extension of the model consisted of a discourse component that implemented discourse saliency as memory activation that is affected by the referents' current discourse prominence. The aim of this model was to explain the acquisition of subject pronouns in Dutch. Dutch-speaking children up to 7 years old show non-adult-like interpretation and production of subject pronouns: In their interpretations, they do not seem to use the discourse status of referents in the local discourse as a cue for

subject pronoun interpretation. In their productions, they overuse pronouns, even for reference to an individual that is not the discourse topic. On the basis of our extended cognitive model, we predicted that children's non-adult-like pronoun processing may have two different causes. First, children may not have sufficient processing speed for taking into account the perspective of a hypothetical conversational partner in their interpretation and production. Second, children may not have sufficient WM capacity to process discourse structure in an adult-like way.

In **Chapter 4**, I presented the results of a dual-task experiment with adults providing support for the prediction that the mature use of discourse structure and the linguistic constraints pertaining to this discourse structure require sufficient WM capacity. If WM load in adults is increased due to a difficult secondary task, less WM capacity is available for sentence processing. In such a dual-task situation, adult participants were less likely to use the discourse status of referents to interpret subject pronouns. Thus, they started to resemble children in their linguistic behavior.

In a follow-up study described in **Chapter 5** we measured EEG during a similar dual task to test whether participants consider alternative referents for the pronoun during *on-line* pronoun resolution. The results of this ERP study show that the preceding discourse already resolved the ambiguity in the discourse before the pronoun was read: Effects of WM load and topic shift were found at the introduction of a second referent in the prominent subject position, which causes a shift of topic. Effects of WM load and topic shift were also present on reading the pronoun, but considerably reduced. The results indicate that the relative discourse saliency of referents constrains the choice for a referent during on-line pronoun processing, although the discourse ambiguity may play a role in off-line pronoun resolution.

In this thesis, I proposed that subject pronouns are produced and interpreted using the same underlying mechanism as for object pronouns: the cognitive model explaining object pronoun acquisition was extended with a discourse processing component to explain subject pronoun acquisition. If object pronouns and subject pronouns make use of similar underlying mechanisms, combining the two cognitive models would yield the prediction that the on-line processing of object pronouns, but not necessarily the outcome of the process, is influenced by the structure of the preceding discourse in a similar way as the discourse influences subject pronoun processing. **Chapter 6** described an experiment testing this prediction. We measured pupil dilation to investigate whether discourse may change the processing load during object pronoun interpretation. Our results indicate that the introduction order of referents, which affects the referents' relative discourse saliency, has a very early influence on object pronoun resolution.

This thesis has yielded a cognitively motivated account of the effects of processing speed and WM capacity on pronoun processing. This account has implications for theories of language acquisition, for models of sentence processing and for the communication deficits associated with cognitive disorders.

2.1 Cognitive factors in pronoun processing

Our findings support the predictions of our cognitive model that both speed of processing and WM capacity play an important role in pronoun processing, albeit in different ways. Insufficient speed of processing interacts with the use of grammatical constraints in pronoun processing, and may impede the process of perspective taking, a Theory of Mind-like process. Insufficient WM capacity, on the other hand, affects discourse processing, as it hinders the mature use of local aspects of the discourse structure such as the grammatical role of referents in the discourse. In this thesis, we did not investigate the interaction between WM capacity and processing speed. In ACT-R, processing speed is influenced by WM capacity, because higher WM capacity increases the activation of referents and concepts in declarative memory, resulting in shorter retrieval times. However, these increases in processing speed due to higher WM capacity are outperformed by the increase in processing speed due to ACT-R's learning mechanism of production compilation. This learning process incorporates all declarative knowledge in highly specialized production rules, so that retrieval of declarative knowledge is unnecessary (Taatgen & Anderson, 2002, see Chapter 2 and 3). Although WM capacity and speed of processing may interact, they affect the acquisition, use and interpretations of pronouns differently.

3 Theoretical implications

3.1 Implications for language acquisition research

Our cognitively motivated account of pronoun processing has a number of implications for language acquisition research. We proceeded from the assumption that 4 to 7-year-old Dutch-speaking children already possess mature knowledge of the constraints of the grammar pertaining to pronoun use, and their relative ranking. However, they still make systematic errors in their interpretation (of object pronouns, see Koster, 1993; Philip & Coopmans, 1996; Spenader et al., 2009) and their production (of subject pronouns, see Wubs et al., 2009). We argued in previous chapters that these errors disappear if children are capable of perspective taking and mature discourse processing.

Importantly, our model predicts that sufficient processing speed and sufficient WM capacity are not necessarily acquired around the same time, because these cognitive factors are acquired through different underlying mechanisms. In our cognitive model, processing speed is crucially dependent on the input frequency of the pronouns. This is because processing speed increases as a result of the repeated use of production rules in ACT-R, due to ACT-R's learning mechanism (cf. Taatgen & Anderson, 2002). On the other hand, in ACT-R spreading activation reflecting the amount of WM capacity available for maintaining task relevant information ready to use is independent of experience and frequency. This is in line with other capacity accounts of working memory, who assume WM capacity to be acquired through maturation (a.o., Just & Carpenter, 1992; Daneman & Carpenter, 1980).

So children must acquire sufficient WM capacity to process pronouns in an adult-like way. Interestingly, even in adults the amount of available WM capacity influences pronoun pro-

cessing. In Chapter 4 and 5, we discussed our finding that a secondary task with a high rather than low WM load, which results in less available WM capacity for pronoun processing, alters the way adults interpret short stories containing a pronoun. This finding goes against the frequency-based account of working memory of MacDonald and Christiansen (2002), who propose that individual differences in WM capacity emerge from an interaction of biological factors and language experience. Although we agree that language experience plays an important role in sentence processing, for example in the acquisition of perspective taking, language experience cannot account for the change in interpretation within the same individual as a result of a higher WM load.

In addition to sufficient WM capacity, children need to be able to take into account the perspective of their conversational partner. As speakers, they need to be able to take into account the perspective of the listener, and as listeners, they need to be able to take into account the perspective of the speaker. With sufficient processing speed, these two processes of considering one's own perspective and subsequently considering the perspective of one's conversational partner can be completed within a reasonable time. Processing speed increases through linguistic experience, which is dependent on the frequencies of the forms and meanings in the language. The complex interaction between the frequency of a form or meaning in child-directed speech and WM capacity, which are both relevant for mature sentence processing, may be responsible for the large individual differences in the age of acquisition of pronoun interpretation and production. For example, some children may already interpret object pronouns correctly at the age of 4, whereas other typically-developing children continue to make interpretation errors until the age of 6.

The acquisition of referential skills is not an easy job. Children need to acquire the relevant grammatical constraints and their relative weights, they need to develop their perspective taking abilities, they need to acquire sufficient WM capacity for discourse processing, and they need to gain sufficient experience with pronoun processing to be able to use perspective taking during on-line pronoun production and interpretation. A basic assumption within our implementation is that children already start out with the general ability to take into account the opposite perspective. The process of perspective taking used for pronoun interpretation and production is a specialization of the general process of perspective taking, which may also be applied in other (linguistic or non-linguistic) tasks. Because adult-like use of perspective taking is dependent on frequency of the forms and meanings in the language, the time to acquire adult-like use of perspective taking is expected to be different for the various linguistic phenomena that require perspective taking.

3.2 Implications for models of sentence processing

In this thesis, I propose a unified account of pronoun processing that explains the processing of object pronouns and subject pronouns using the same underlying mechanism. Traditionally, *subject pronouns* are studied in the domain of pragmatics, where the effects of the preceding discourse are investigated. *Object pronouns*, on the other hand, are studied in the domain of syntax (although recent years have seen a move away from purely syntactic accounts of ob-

ject pronouns to accounts at the syntax-pragmatics interface, see, e.g., Reinhart, 2006), where the grammatical principles guiding sentence generation and interpretation are investigated. In this thesis, I argue that the same underlying mechanism can be used to explain the use of pronouns in both linguistic environments. Obviously, grammatical principles play a less prominent role in the processing and use of subject pronouns than in the processing and use of object pronouns. Conversely, the preceding discourse plays a less prominent role in the processing and use of object pronouns than in the processing and use of subject pronouns. In languages such as English and Dutch, on which I focused in this thesis, subject pronouns are less constrained by grammatical principles, because they tend to appear relatively early in the sentence and occur in a syntactically prominent position. Object pronouns are less influenced by the preceding discourse, because they tend to occur later in the sentence and the number of possible interpretations becomes smaller as the sentence unfolds. However, we still can find influences of the preceding linguistic discourse on the on-line interpretation of object pronouns, as our pupil dilation study in Chapter 6 shows. This provides evidence for the lack of a fundamental distinction between sentence-based processes and discourse-based processes in pronominal reference.

The implementation of pronoun processing as proposed in this thesis is in line with constraint-based accounts of sentence processing (a.o., MacDonald et al., 1994; Trueswell, Tanenhaus, & Garnsey, 1994; Tanenhaus & Trueswell, 1995, see Van Gompel & Pickering, 2007, for review), which assume that all available information may influence sentence processing simultaneously. In the cognitive model, discourse factors have an immediate influence on pronoun processing, because they may change the saliency of referents in the discourse representation. As discourse saliency is implemented in the model as activation in memory, discourse saliency reflects the referent's local relevance in the current context, in addition to the global relevance of the referent based on its history of mentioning. The higher the referent's activation in memory, the more likely it is that the model will retrieve this referent from memory. Thus, the linguistic discourse functions as a gradient filter that makes particular information easier to retrieve. As a result, this information is more likely to be used at a later moment in discourse, as it is more readily available during language processing. This anticipatory effect of discourse is in line with memory-based accounts of sentence processing and production (a.o., R. L. Lewis & Vasishth, 2005; Reitter et al., 2011) and memory-based accounts of referent processing (a.o., Foraker & McElree, 2007; Ariel, 1990; Arnold, 1998, see a.o., Otten & Van Berkum, 2009, for empirical support).

Our cognitive model uses perspective taking to validate whether the mapping between a form and the interpretation of this form selected by a listener is also optimal from the perspective of a speaker. The process of perspective taking in the model is particularly important for resolving ambiguities. This thesis focused on pronoun processing, but an interesting question is whether this mechanism may be generalized to other potential ambiguities in sentence processing.

3.3 Implications for accounts of communication deficits

The account of pronoun processing proposed in this thesis may also be able to explain the difficulties with pronoun processing in individuals with language disorders or cognitive disorders. As sufficient processing speed is necessary for taking into account the perspective of the conversational partner, and sufficient working memory (WM) capacity is necessary for discourse processing, individuals who have a low WM capacity (such as elderly people whose memory resources are decreasing due to cognitive decline), individuals who have insufficient processing speed (such as patients with Broca's aphasia, cf. Burkhardt, Avrutin, Piñango, & Ruigendijk, 2008) or difficulties with perspective taking (e.g., individuals with an autistic spectrum disorder) may experience difficulties with pronoun processing. However, these factors do not all cause the same problems with pronoun processing. For example, individuals with insufficient WM capacity but no difficulties with perspective taking may produce subject pronouns that are ambiguous in the context of use, because they have a different discourse representation. Indeed, this was found when investigating the referential skills of elderly adults in narrative discourse (Hendriks et al., 2008). The elderly produced more pronouns for reference to a referent that was not the topic of the discourse than young adults, and this overproduction of pronouns was correlated with lower scores on a WM task (Hendriks et al., 2008). People with a low WM capacity, but no difficulties with perspective taking, are not expected to show differences with the interpretation of object pronouns, which is less influenced by WM capacity. Individuals who have difficulties with perspective taking, because of a slow processing speed or because they have limited Theory of Mind skills, are expected to show difficulties with the interpretation of object pronouns and subject pronouns, as both require perspective taking.

4 Implementation issues

Implementing linguistic theories within a cognitive model yielded specific and testable predictions. In addition, it allowed us to specify the assumptions and implications of the linguistic theories on which we based our model. However, the integration of different theories sometimes resulted in implementational choices that need further investigation.

In the cognitive model, I combined linguistic constraints from Optimality Theory with discourse processes based on ACT-R's memory principles. Discourse saliency was implemented as memory activation (cf. Foraker & McElree, 2007). Thus, discourse factors have a gradual effect. The linguistic constraints, however, assume that there is only one discourse topic (cf. Grosz et al., 1995; Gordon et al., 1993; Beaver, 2004). This step from a gradual pattern of activation of discourse referents to a discrete choice of one of these referents is achieved by retrieving the referent with the highest activation at that point in the discourse as the discourse topic. Because some of the linguistic constraints pertain to the current discourse topic, identification of the discourse topic precedes the application of linguistic constraints, As a consequence of this implementational choice, grammatical processing follows discourse processing. It would be interesting to investigate whether speakers indeed have to choose the

current topic for pronoun use, or whether a more gradual notion of topic is used, reflecting the likelihood of a referent being the topic. To address this question, however, the grammatical principles need to be redefined in the cognitive model so that they base the choice for a referring expression on the likelihood of a referent being the topic instead of retrieving the topic for comparison. As not having to explicitly retrieve the topic saves costly time during on-line production, this process may be more realistic (see also Foraker & McElree, 2007, for a similar discussion). Note however that this alternative implementation would not change the results described in this thesis.

Another implementational choice that may affect the performance of the model is the assumption that discourse processing is similar for speakers and listeners, as discussed in Chapter 3. This would imply that a speaker and a listener base their selected form and interpretation on the same discourse representation. However, it may be the case that speaking about a referent makes the referent more salient in the discourse than listening to this reference: the process of choosing a referring expression for a particular referent may increase the saliency more than interpreting a referring expression, or vice versa. Because this issue falls outside the scope of this thesis, I chose to simplify the implementation of pronoun processing by assuming that producing referential expressions and interpreting referential expressions result in similar discourse representations.

The final implementational choice I want to address is the implementation of WM capacity. I based my implementation on the work of Daily et al. (2001), who modeled individual differences in ACT-R by manipulating the amount of spreading activation from task-relevant information to associated information in declarative memory. In this account, working memory in a certain sense is the activation of task-relevant information in memory. It is important to realize, however, that ACT-R does not assume a single mechanism to capture working memory. For example, another aspect of working memory is realized by ACT-R's assumption that only one piece of information can be retrieved and manipulated at a time for each module (Borst et al., 2010). More research is needed to identify the effects of other aspect of WM capacity on pronoun processing.

5 Conclusions

In every conversation we encounter pronouns, and we need to decide what the antecedents of these pronouns are and whether we can use a pronoun to refer to a specific referent. We are generally not consciously aware that during pronoun resolution we integrate information from the preceding linguistic discourse, the grammar and our conversational partner's perspective. Furthermore, we may also other sources of information that were not investigated in this thesis, such as the visual context, prosodic information and world knowledge. Limations of our cognitive system, such as insufficient speed of processing, insufficient WM capacity or difficulties with perspective taking, may hinder this process, causing difficulties for language processing and communication. Thus, this thesis addresses a relevant topic, of which the insights may have various applications.

Investigating the complex interaction between cognitive factors and pronoun processing

requires an interdisciplinary approach. Previous research, focusing on the grammatical aspects of pronoun processing or the discourse factors influencing referential choice has yielded new insights and theories. However, as pronoun processing involves combining information from different sources and even modalities, our methodology may be a valuable addition to this research. It combines insights from different domains, resulting in an integrated account of pronoun acquisition, use and interpretation. Cognitive modeling is a useful method for studying complex interactions between domain-specific (e.g., linguistic) knowledge and the limitations of the cognitive system, by using simulations within a unified theory of cognition. In addition, computational modeling helps to specify the assumptions of the theories that are modeled, and allows for the generation of cognitively plausible and testable predictions that can be investigated with behavioral and psychophysiological measures (e.g., ERP and pupil dilation).

To summarize, this thesis presented a novel account of pronoun processing that integrates theories of grammar and discourse, and explains how cognitive factors such as speed of processing and WM capacity interact with discourse and grammar in pronoun production and interpretation. In addition, this integrated account may provide new insights in theories of language acquisition, models of sentence processing, and accounts of communication deficits.

References

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

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036-1060.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396-408.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, *24*(1), 65-87.

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

Arnold, J. E. (1998). *Reference form and discourse patterns*. Ph.D. thesis.

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and cognitive processes*, *23*(4), 495-527.

Arnold, J. E., Bennetto, L., & Diehl, J. J. (2009). Reference production in young speakers with and without autism: Effects of discourse status and processing constraints. *Cognition*, *110*(2), 131-146.

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. C. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, *22*(4), 527-565.

Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, *56*(4), 521-536.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12-28.

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.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*(1), 1-22.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'Theory of Mind'? *Cognition*, *21*(1), 37-46.

Bates, D. M. (2005). Fitting linear mixed models in R. *R News*, *5*, 27-30.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276-292.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., p. 142-162). New York, NY: Cambridge University Press.

Beaver, D. (2004). The optimization of discourse anaphora. *Linguistics and Philosophy*, *27*(1), 3-56.

Bloom, P., Barss, A., Nicol, J., & Conway, L. (1994). Children's knowledge of binding and coreference: Evidence from spontaneous speech. *Language(Baltimore)*, *70*(1), 53-71.

Blutner, R. (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics*, *17*(3), 189-216.

Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, *32*(1), 45-86.

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.

Bosch, V., Mecklinger, A., & Friederici, A. D. (2001). Slow cortical potentials during retention of object, spatial, and verbal information. *Cognitive Brain Research*, *10*, 219-237.

Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive Processes*, *10*(2), 137-167.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*(2), 274-291.

Budiu, R., & Anderson, J. R. (2004). Interpretation-based processing: A unified theory of semantic sentence comprehension. *Cognitive Science*, *28*(1), 1-44.

Burkhardt, P., Avrutin, S., Piñango, M. M., & Ruigendijk, E. (2008). Slower-than-normal syntactic processing in agrammatic Broca's aphasia: Evidence from Dutch. *Journal of Neurolinguistics*, *21*, 120-137.

Burzio, L. (1998). Anaphora and soft constraints. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, & D. Pesetsky (Eds.), *Is the best good enough? Optimality and competition in syntax* (p. 93-114). Cambridge, MA: MIT Press.

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*, *56*, 103-128.

Campana, E., Tanenhaus, M. K., Allen, J. F., & Remington, R. (2011). Natural discourse reference generation reduces cognitive load in spoken systems. *Natural Language Engineering*, *17*(3), 311-329.

Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behavior*, *16*, 601-609.

Case, R. (1987). The structure and process of intellectual development. *International Journal of Psychology*, *22*(5), 571-607.

Chambers, C. G., & Smyth, R. (1998). Structural parallelism and discourse coherence: A test of Centering Theory. *Journal of Memory and Language*, *39*(4), 593-608.

Chien, Y. C., & Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition*, *1*(3), 225-295.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding: the Pisa lectures*. Dordrecht, The Netherlands: Foris Publications.

Clackson, K., Felser, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in english: Evidence from eye-movements during listening. *Journal of Memory and Language*, *65*, 128-144.

Clark, H. H. (1996). *Using language.* Cambridge University Press.

Clark, H. H., & Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Memory and Cognition*, *7*(1), 35-41.

Clifton, C., Kennison, S., & Albrecht, J. (1997). Reading the words her, his, him: Implications for parsing principles based on frequency and on structure. *Journal of Memory and Language*, *36*(2), 276-292.

Crawley, R., Stevenson, R., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, *19*(4), 245-264.

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.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450-466.

De Villiers, J., Cahillane, J., & Altreuter, E. (2006). What can production reveal about Principle B? In K. Deen, J. Nomura, B. Schulz, & B. Schwartz (Eds.), *The Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition-North America* (Vol. 1, p. 89-100). Honolulu, HI.: University of Connecticut Occasional Papers in Linguistics 4. (The Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition-North America, University of Connecticut Occasional Papers in Linguistics)

Dijkstra, T., & De Smedt, K. (1996). Computer models in psycholinguistics: An introduction. *Computational psycholinguistics*, 3-23.

Ehrlich, K. (1980). Comprehension of pronouns. *Quarterly Journal of Experimental Psychology*, *32*(2), 247-255.

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior*, *22*(1), 75-87.

Elbourne, P. (2005). On the acquisition of principle b. *Linguistic Inquiry*, *36*(3), 333-365.

Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *The Quarterly Journal of Experimental Psychology*, *63*(4), 639-645.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211-245.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and

lexical production. *Cognitive Psychology*, *40*(4), 296-340.
Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, *56*(3), 357-383.
Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance xii: The psychology of reading* (p. 559-586). Hillsdale, NJ: Erlbaum.
Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, *27*(6), 699-717.
Givón, T. (1983). Topic and continuity in discourse: An introduction. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (p. 1-42). Amsterdam: John Benjamins Publishing.
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.
Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3: Speech acts, p. 41-58). New York: Academic Press.
Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274-279.
Grimshaw, J., & Rosen, S. T. (1990). Knowledge and obedience: The developmental status of the Binding Theory. *Linguistic inquiry*, *21*(2), 187-222.
Grober, E. H., Beardsley, W., & Caramazza, A. (1978). Parallel function strategy in pronoun assignment. *Cognition*, *6*(2), 117-133.
Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, *21*(2), 203-225.
Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274-307.
Gundel, J. K., Ntiletheos, D., & Kowalski, M. (2007). Children's use of referring expressions: Some implications for theory of mind. *ZAS Papers in Linguistics*, *48*, 1-21.
Hale, C. M., & Tager-Flusberg, H. (2005). Social communication in children with autism. *Autism*, *9*(2), 157.
Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science*, *16*(5), 362-366.
Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831-836.
Hendriks, P., & de Hoop, H. (2001). Optimality theoretic semantics. *Linguistics and Philosophy*, *24*(1), 1-32.
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.
Hendriks, P., & Spenader, J. (2006). When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition*, *13*(4), 319-348.
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.
Hickmann, M., & Hendriks, H. (1999). Cohesion and anaphora in children's narratives: A comparison of English, French, German, and Mandarin Chinese. *Journal of Child Language*,

26(2), 419-452.

Hoeks, B., & Levelt, W. J. M. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods*, *25*(1), 16-26.

Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, *19*(4), 286-300.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91-117.

Hyönä, J., Tommola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology*, *48*(3), 598-612.

Jakubowicz, C. (1984). On markedness and binding principles. *Proceedings of the North Eastern Linguistics Society*, *14*, 154-182. (Proceedings of the North Eastern Linguistics Society)

Janisse, M. (1977). *Pupillometry.* Washington, D.C.: Hemisphere.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122-149.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimensionof thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, *47*(2), 310-339.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology*, *111*(2), 228-238.

Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, and Behavioral Neuroscience*, *7*(3), 153-191.

Kaiser, E., & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, *23*(5), 709-748.

Karmiloff-Smith, A. (1985). Language and cognitive processes from a developmental perspective. *Language and Cognitive Processes*, *1*(1), 61-85.

Kempen, G., Olsthoorn, N., & Sprenger, S. (2011). Grammatical workspace sharing during language production and language comprehension: Evidence from grammatical multitasking. *Language and Cognitive Processes*.

Kennison, S. M. (2003). Comprehending the pronouns *her*, *him*, and *his*: Implications for theories of referential processing. *Journal of Memory and Language*, *49*(3), 335-352.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, *7*, 46-50.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*(3), 376-395.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New york, NY: Cambridge University Press.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*, 95-127.

Koster, C. (1993). *Errors in anaphora acquisition.* Unpublished doctoral dissertation.

Koster, C., Hoeks, J., & Hendriks, P. (2011). Comprehension and production of subject pronouns: Evidence for the asymmetry of grammar. In A. Grimm, A. Müller, C. Hamann, & E. Ruigendijk (Eds.), *Production-comprehension asymmetries in child language* (p. 99-122). Berlin: De Gruyter.

Kutas, M., & Federmeier, K. D. (2007). Event-related brain potential (ERP) studies of sentence processing. In M. G. Gaskell (Ed.), *Oxford handbook of psycholinguistics.* Oxford University Press.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain-potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205.

Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161-163.

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified ii (1994-2005). In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics.* Elsevier Ltd.

Ledoux, K., Gordon, P. C., Camblin, C. C., & Swaab, T. Y. (2007). Coreference and lexical repetition: Mechanisms of discourse integration. *Memory and Cognition*, *35*(4), 801-815.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375-419.

Lewis, S., Chow, W., & Phillips, C. (2012, 2012). *Structural constraints on pronoun resolution: Distinguishing early and late sensitivity to ilicit antecedents.*

Love, T., Walenski, M., & Swinney, D. (2009). Slowed speech input has a differential impact on on-line and off-line processing in children's comprehension of pronouns. *Journal of Psycholinguistic Research*, *38*(3), 285-304.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*(1), 35-54.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676-703.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2009). Pronoun co-referencing errors: Challenges for generativist and usage-based accounts. *Cognitive Linguistics*, *20*(3), 599-626.

McClelland, J. L. (1995). A connectionist perspective on knowledge and development. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 157–204). Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: I. basic mechanisms. *Psychological Review*, *104*(1), 3-65.

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* (p. 165-170). Bamberg: Universitäts-Verlag Bamberg.

Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the reso-

lution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language*, *33*(2), 175-202.

Montgomery, J. W. (2004). Sentence comprehension in children with specific language impairment: effects of input rate and phonological working memory. *International Journal of Language and Communication Disorders*, *39*(1), 115-133.

Münte, T. F., Heinze, H.-J., & Mangun, G. R. (1993). Dissociation of brain activity related to syntactic and semantic aspects of language. *Journal of Cognitive Neuroscience*, *5*(3), 335-344.

Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, *395*(71-73).

Myachykov, A., Thompson, D., Scheepers, C., & Garrod, S. (2011). Visual attention and structural choice in sentence production across languages. *Language and Linguistics Compass*, *5*(2), 95-107.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*(1), 5-19.

Nieuwland, M. S., & van Berkum, J. J. A. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, *1118*(1), 155-167.

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*, 544-569.

Onishi, K., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255-258.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 1-9.

Otten, M., & van Berkum, J. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain research*, *1291*, 92-101.

Philip, W., & Coopmans, P. (1996). The double Dutch delay of Principle B effect. In *Annual boston university conference on language development* (Vol. 20).

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*, 407-428.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515-526.

Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science*, *275*(5306), 1604.

Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar.* Malden, MA: Blackwell.

Reinhart, T. (2006). *Interface strategies: Optimal and costly computations*. Cambridge, MA: MIT Press.

Reitter, D. (2008). *Context effects in language production: Models of syntactic priming in dialogue corpora*. Ph.D. thesis.

Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive Science*, *35*.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, *115*(1), 101.

Scheepers, C., & Crocker, M. W. (2004). Constituent order priming from reading to listening: A visual-world study. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs, and beyond* (p. 167-186). New York, NY: Psychology Press.

Shintel, H., & Keysar, B. (2009). Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science*, *1*(2), 260-273.

Small, J. A., Andersen, E. S., & Kempler, D. (1997). Effects of working memory capacity on understanding rate-altered speech. *Aging, Neuropsychology, and Cognition*, *4*(2), 126 - 139.

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry*, *27*(4), 720-731.

Song, H.-J., & Fisher, C. (2005). Who's "she"? discourse prominence influences preschoolers' comprehension of pronouns. *Journal of Memory and Language*, *52*(1), 29-57.

Song, H.-J., & Fisher, C. (2007). Discourse prominence effects on 2.5-year-old children's interpretation of pronouns. *Lingua*, *117*(11), 1959-1987.

Spenader, J., Smits, E. J., & Hendriks, P. (2009). Coherent discourse solves the Pronoun Interpretation Problem. *Journal of Child Language*, *36*(1), 23-52.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye-movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*(4), 447-481.

Sprenger, S. A., Levelt, W. J. M., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, *54*(2), 161-184.

St. George, M., Mannes, S., & Hoffman, J. E. (1994). Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience*, *6*(1), 70-83.

Stevenson, R., Crawley, R., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, *9*(4), 519-548.

Stevenson, S., & Smolensky, P. (2006). Optimality in sentence processing. In P. Smolensky & G. Legendre (Eds.), *The harmonic mind: From neural computation to optimality-theoretic grammar* (Vols. 2, Linguistic and philosophical implications). Cambridge, MA: MIT Press.

Streb, J., Hennighausen, E., & Rösler, F. (2004). Different anaphoric expressions are investigated by event-related brain potentials. *Journal of Psycholinguistic Research*, *33*(3), 175-201.

Streb, J., Rösler, F., & Hennighausen, E. (1999). Event-related responses to pronoun and proper name anaphors in parallel and nonparallel discourse structures. *Brain and Language*, *70*, 273-286.

Swaab, T. Y., Camblin, C. C., & Gordon, P. C. (2004). Electrophysiological evidence for reversed lexical repetition effects in language processing. *Journal of Cognitive Neuroscience*, *16*(5), 715-726.

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.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. *Speech, language, and communication*, *18*, 217–262.

Tesar, B., & Smolensky, P. (1998). Learnability in optimality theory. *Linguistic Inquiry*, *29*(2), 229–268.

Thornton, R., & Wexler, K. (1999). *Principle B, VP ellipsis, and interpretation in child grammar*. MIT Press.

Tily, H., & Piantadosi, S. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. In K. van Deemter, A. Gatt, R. van Gompel, & E. Krahmer (Eds.), *Proceedings of the pre-cogsci 2009 workshop on production of referring expressions: Bridging the gap between computational and empirical approaches to reference.*

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and erp study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication.* (p. 151-173). London: The Continuum International Publishing Group.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285-285.

van Berkum, J. J. A. (2008). Understanding sentences in context. *Current Directions in Psychological Science*, *17*(6), 376-380.

van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657-671.

van Berkum, J. J. A., Koorneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, *1146*, 158-171.

van Gompel, R. P. G., & Pickering, M. J. (2007). Syntactic parsing. In M. G. Gaskell (Ed.), *The oxford handbook of psycholinguistics.* (p. 289-307). Oxford, UK: Oxford University Press.

van Maanen, L., & van Rijn, H. (2007). An accumulator model of semantic interference. *Cognitive Systems Research*, *8*(3), 174-181.

van Maanen, L., van Rijn, H., & Borst, J. P. (2009). Stroop and picture-word interference are two sides of the same coin. *Psychonomic Bulletin and Review*, *16*(6), 987-999.

Van Petten, C. K. (1995). Words and sentences: Event-related brain potential measures. *Psychophysiology*, *32*, 511-525.

Van Petten, C. K., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, *18*(4), 380-393.

van Rij, J., Hendriks, P., Spenader, J., & van Rijn, H. (2009*a*). From group results to individual patterns in pronoun comprehension. In J. Chandlee, M. Franchini, S. Lord, & M. Rheiner (Eds.), *Proceedings of the 33rd annual Boston University Conference on Language Development* (Vol. 2, p. 563-574). Somerville, MA: Cascadilla Press.

van Rij, J., Hendriks, P., Spenader, J., & van Rijn, H. (2009*b*). Modeling the selective effects of slowed-down speech in pronoun comprehension. In J. Crawford, K. Otaki, & M. Taka-

hashi (Eds.), *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008)* (p. 291-302). Somerville, MA: Cascadilla Proceedings Project.

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. (Chapter 2 of this thesis)

van Rij, J., van Rijn, H., & Hendriks, P. (2011). WM load influences the interpretation of referring expressions. In *Proceedings of the 2nd workshop on cognitive modeling and computational linguistics* (p. 67-75). Portland, OR: Association for Computational Linguistics.

van Rij, J., van Rijn, H., & Hendriks, P. (submitted*a*). A cognitive model of the acquisition and use of referring expressions: Effects of working memory and perspective taking. Manuscript, University of Groningen, October 2011. (Chapter 3 of this thesis)

van Rij, J., van Rijn, H., & Hendriks, P. (submitted*b*). How WM load influences linguistic processing in adults: A computational model of pronoun interpretation in discourse. Manuscript, University of Groningen, February 2012. (Chapter 4 of this thesis)

van Rijn, H., & Anderson, J. R. (2003). Modeling lexical decision as ordinary retrieval. In F. Detje, D. Dörner, & H. Schaub (Eds.), *Proceedings of the fifth international conference on cognitive modeling* (p. 207-212). Bamberg: Universitätsverlag.

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.

Wardlow Lane, L., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1466-1481.

Weismer, S. E., & Hesketh, L. J. (1996). Lexical learning by children with specific language impairment effects of linguistic input presented at varying speaking rates. *Journal of Speech, Language and Hearing Research*, *39*(1), 177-190.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta analysis of theory of mind development: The truth about false belief. *Child development*, *72*(3), 655-684.

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, *6*(9), e23613.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Chapman and Hall/CRC.

Wubs, E., Hendriks, P., Hoeks, J., & Koster, C. (2009). Tell me a story! Children's capacity for topic shift. In J. Crawford, K. Otaki, & M. Takahashi (Eds.), *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008)* (p. 313-324). Somerville, MA: Cascadilla Press.

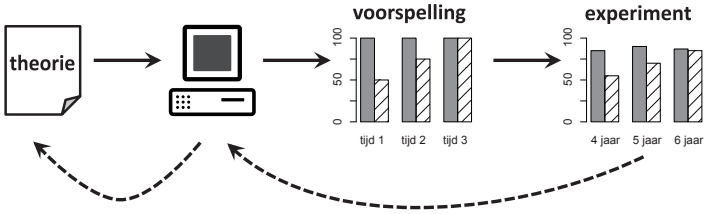
Zellin, M., Pannekamp, A., Toepel, U., & van der Meer, E. (2011). In the eye of the listener: Pupil dilation elucidates discourse processing. *International Journal of Psychophysiology*, *81*(3), 133-141.

Nederlandse samenvatting

De uitspraak *Gisteren sprak James met Rob. Hij bekende de diefstal.* kan betekenen dat James de diefstal bekende, of dat Rob dat deed. Deze zin is *ambigu*, dat wil zeggen dat er twee mogelijke betekenissen zijn, omdat het persoonlijke voornaamwoord *hij* naar James kan verwijzen of naar Rob. Persoonlijke voornaamwoorden zoals *hij*, *zij*, *hem* en *haar* hebben geen vaste betekenis, maar hun betekenis is afhankelijk van de context. Persoonlijke voornaamwoorden zijn cruciaal voor taalbegrip: als je een andere interpretatie voor een voornaamwoord kiest, dan krijgt de zin een andere betekenis. Vervolgens begrijp je het verhaal of de conversatie waarschijnlijk ook verkeerd. Daarom is er veel onderzoek gedaan naar de verschillende taalkundige factoren die een rol spelen bij de interpretatie en het gebruik van voornaamwoorden, zoals grammaticale principes en de structuur van de voorafgaande zinnen.

Het is opvallend dat volwassenen meestal geen enkele moeite hebben met het interpreteren van een mogelijk ambigu voornaamwoord, maar en vaak meteen begrijpen wat de spreker bedoelde te zeggen. Kinderen hebben hier meer moeite mee, want ze interpreteren persoonlijke voornaamwoorden soms anders dan de spreker bedoelde. Ze verwerven pas relatief laat in de taalontwikkeling een volwassen interpretatie van voornaamwoorden, meestal niet voordat ze zes jaar zijn. In dit onderzoek is onderzocht waarom persoonlijke voornaamwoorden in talen zoals het Nederlands en Engels zo moeilijk zijn om te leren, en welke algemene cognitieve factoren van invloed zijn op een volwassen verwerking van voornaamwoorden.

Om deze vragen te beantwoorden hebben we een combinatie van onderzoeksmethoden gebruikt, waaronder computersimulaties en experimenten met kinderen en volwassenen. Figuur 1 toont schematisch de methode die we hebben toegepast.



Figuur 1: Een schematische weergave van de onderzoeksmethode die is toegepast in dit onderzoek.

De computermodellen werden ontwikkeld om precieze en cognitief plausibele voorspellingen te doen over de interpretatie en het gebruik van persoonlijke voornaamwoorden. Deze

voorspellingen zijn zo precies mogelijk geformuleerd zodat ze getest kunnen worden door middel van experimenten. De resultaten van de experimenten kunnen worden gebruikt om te evalueren in hoeverre de computermodellen correct zijn (zie de onderbroken lijn rechts in Figuur 1). Ook kunnen de resultaten nieuwe informatie opleveren voor taalkundige theorieën (zie de onderbroken lijn links in Figuur 1).

In de volgende paragrafen worden eerst de computermodellen beschreven en de voorspellingen die volgen uit de simulaties. Daarna worden de verschillende experimenten beschreven en wordt er een kort overzicht gegeven van de resultaten. Aan het eind van deze samenvatting volgen de conclusies van dit onderzoek.

Computermodellen van taalverwerving

In deze sectie worden de computermodellen beschreven. Ze zijn gebaseerd op verschillende taalkundige theorieën en simuleren hoe kinderen persoonlijke voornaamwoorden leren. De computermodellen zijn geïmplementeerd in de cognitieve architectuur ACT-R (Anderson, 2007).

Cognitieve architectuur. Een cognitieve architectuur is een theorie over hoe het cognitieve systeem werkt. Cognitieve architecturen, bijvoorbeeld ACT-R, proberen menselijke cognitie te verklaren in allerlei mogelijke verschillende taken variërend van autorijden tot het begrijpen van een zin. Ze bevatten precieze aannames over de werking van het menselijke cognitieve systeem, zoals bijvoorbeeld hoe lang het duurt om op een knop te drukken of hoeveel tijd het kost om een feit uit het geheugen op te halen. De aannames zijn gebaseerd op de resultaten van veel psychologische en neurofysiologische experimenten. Met behulp van deze aannames kunnen we in ACT-R simuleren hoe kinderen en volwassenen persoonlijke voornaamwoorden, zoals *hij* en *hem*, begrijpen in een specifieke taak. Deze simulaties kunnen vervolgens worden gebruikt om te onderzoeken hoe algemene cognitieve factoren, zoals verwerkingssnelheid en werkgeheugencapaciteit, een rol spelen in het verwerken van voornaamwoorden.

Hoofdstuk 1 introduceert de cognitieve architectuur ACT-R en geeft een overzicht van verschillende aannames van ACT-R die onze implementaties beïnvloeden. Hieronder worden kort de twee modellen beschreven die zijn ontwikkeld in het kader van dit onderzoek. Deze modellen zijn in meer detail beschreven in hoofdstuk 2 en hoofdstuk 3.

Model 1: verwerving van voornaamwoorden in objectpositie

Hoofdstuk 2 beschrijft het eerste computermodel dat is ontwikkeld voor dit onderzoek. Dit model simuleert hoe kinderen woorden zoals *hem* en *haar* leren te begrijpen. *Hem* en *haar* zijn persoonlijke voornaamwoorden die als direct object (lijdend voorwerp) kunnen voorkomen. Denk bijvoorbeeld aan de uitspraak *Hier zie je een pinguïn en een schaap. De pinguïn slaat hem met een pan.* Voor volwassenen kan dit alleen betekenen dat de pinguïn het schaap slaat, want anders zou de spreker wel gezegd hebben dat de pinguïn *zichzelf* slaat. Maar in talen zoals het Engels en het Nederlands blijken kinderen tot ongeveer zeven jaar veel fouten te maken met de interpretatie van dit soort zinnen. Voor hen kan de zin betekenen dat de pinguïn het schaap slaat, maar evengoed dat de pinguïn zichzelf slaat.

In het computermodel hebben we de hypothese geïmplementeerd dat kinderen moeite hebben met voornaamwoorden in objectpositie, omdat ze geen rekening houden met het perspectief van de spreker. Als een luisteraar rekening houdt met de spreker, dan komt hij/zij erachter dat de spreker niet naar de pinguïn verwijst, want dan had de spreker wel het woord *zichzelf* gebruikt in plaats van *hem*. In het model wordt dit proces in twee stappen gesimuleerd. In de eerste stap gebruikt het model de grammaticale principes om te bepalen naar wie het voornaamwoord kan verwijzen. In de tweede stap probeert het model te bedenken welk woord een spreker zou gebruiken om naar die persoon te verwijzen. Op deze manier kan het model erachter komen wie de spreker bedoelde. Hierbij moet opgemerkt worden dat in de computersimulaties het uitvoeren van iedere stap tijd kost.

Simulaties van het model laten zien dat het mogelijk is dat kinderen vanaf een jaar of vier wel al weten dat ze rekening moeten houden met het perspectief van de spreker. Ze hebben echter niet voldoende verwerkingssnelheid om dat te doen tijdens het interpreteren van een zin. Anders gezegd: kinderen hebben alleen maar tijd om de eerste stap uit te voeren, maar op het moment dat ze proberen om ook de tweede stap uit te voeren, moeten ze alweer verder met de volgende woorden om de rest van de zin te kunnen begrijpen.

Op basis van deze simulaties volgt de voorspelling dat voldoende verwerkingssnelheid nodig is om rekening te kunnen houden met het perspectief van de conversatiepartner.

Model 2: verwerving van voornaamwoorden in subjectpositie

Hoofdstuk 3 beschrijft het tweede computermodel. Dit model is een uitbreiding van het vorige model. Het is zo aangepast dat het simuleert hoe kinderen woorden zoals *hij* en *zij* leren begrijpen. *Hij* en *zij* zijn persoonlijke voornaamwoorden die als subject (onderwerp) voorkomen.

Engelstalige en Nederlandstalige kinderen tot ongeveer zeven jaar maken veel fouten met dit soort voornaamwoorden, niet alleen in hun interpretatie, maar ook in hun gebruik van voornaamwoorden. Volwassenen gebruiken een persoonlijk voornaamwoord om te verwijzen naar personen of dingen (referenten) die erg prominent zijn in de context. Wanneer de persoon minder prominent is in de context, dan gebruiken volwassenen vaak een eigennaam (*Rob*) of een zelfstandig naamwoord (*de piraat*). Hierdoor voorkomen ze onduidelijkheid voor de luisteraar. Kinderen gebruiken veel vaker voornaamwoorden. Ze gebruiken zelfs persoonlijke voornaamwoorden voor referenten die niet prominent zijn. In dat geval kunnen luisteraars het voornaamwoord verkeerd interpreteren, namelijk als verwijzing naar een andere, meer prominente, persoon.

Om de interpretatie en de productie van woorden zoals *hij* en *zij* te simuleren, bouwt het tweede computermodel een representatie op in het geheugen van de talige context waarin de zin voorkomt. Dat is nodig om contextuele informatie te gebruiken bij de interpretatie van voornaamwoorden. Daarnaast bewaart het model tijdelijk informatie over de voorgaande zin in het werkgeheugen. Dit geheugenmechanisme zorgt ervoor dat prominente referenten meer kans maken om te worden gekozen als interpretatie van het voornaamwoord dan minder prominente referenten. Maar dit mechanisme werkt alleen bij voldoende werkgeheugen-

capaciteit.

Op basis van de simulaties hebben we de voorspelling gedaan dat er twee oorzaken zijn voor de moeilijkheden die kinderen ervaren bij de interpretatie en het gebruik van voornaamwoorden in subjectpositie: 1) net als bij de verwerving van voornaamwoorden in objectpositie hebben kinderen ook bij voornaamwoorden in subjectpositie onvoldoende verwerkingssnelheid om rekening te kunnen houden met het perspectief van de conversatiepartner en 2) kinderen hebben onvoldoende werkgeheugencapaciteit om de informatie uit de voorgaande zin te gebruiken om te bepalen wat de meest prominente referent is in de context.

Samengevat voorspellen we dus dat voor een volwassen interpretatie en gebruik van voornaamwoorden voldoende verwerkingssnelheid en voldoende werkgeheugencapaciteit nodig zijn. Maar zoals hiervoor is uitgelegd, beïnvloeden deze twee factoren het gebruik van voornaamwoorden op een verschillende manier.

Gedragsexperimenten

Er zijn twee gedragsexperimenten uitgevoerd om deze voorspellingen te testen, namelijk een experiment met kinderen en een experiment met volwassenen.

Experiment 1: Verwerkingssnelheid

Een belangrijke voorspelling op basis van de simulaties is dat kinderen moeite hebben met de interpretatie van voornaamwoorden in objectpositie vanwege een te lage verwerkingssnelheid. Om deze voorspelling te onderzoeken hebben we getest of kinderen tussen de 4 en 6 jaar minder fouten maken wanneer ze meer tijd krijgen voor de verwerking van voornaamwoorden.

In het experiment dat beschreven is in **hoofdstuk 2** kregen de kinderen zinnen in een normaal spraaktempo te horen en zinnen in vertraagde spraak. In het laatste geval hadden ze dus iets meer tijd voor de interpretatie. De resultaten van het experiment laten zien dat kinderen minder fouten maakten met het begrijpen van voornaamwoorden bij het horen van vertraagde spraak dan bij het horen van normale spraak. Maar dit effect gold alleen voor het persoonlijke voornaamwoord *hem*, want vertraagde spraak bleek een nadelig effect te hebben op zinnen met zichzelf.

Met andere woorden, wanneer kinderen vertraagde spraak horen lijken ze meer op volwassenen in het verwerken van voornaamwoorden. Dit ondersteunt de voorspelling van het model dat voor het verwerken van voornaamwoorden voldoende verwerkingssnelheid nodig is.

Experiment 2: Werkgeheugencapaciteit

Een andere voorspelling op basis van de simulaties van het computermodel was dat voor het verwerken van voornaamwoorden in subjectpositie voldoende werkgeheugencapaciteit nodig is. Werkgeheugencapaciteit is nodig om op basis van informatie uit de voorgaande zin te bepalen wat de meest prominente referent in de talige context is.

Hoofdstuk 4 beschrijft een experiment waarin we hebben onderzocht of de interpretatie door volwassenen van voornaamwoorden in subjectpositie verandert bij onvoldoende werkgeheugencapaciteit. Om dat te onderzoeken hebben we de volwassen deelnemers gevraagd om korte verhaaltjes te lezen terwijl ze tegelijkertijd een tweede taak moesten uitvoeren. De tweede taak zorgde ervoor dat de deelnemers meer of minder werkgeheugencapaciteit over hadden om het verhaaltje te begrijpen. De tweede taak was makkelijk of moeilijk: De makkelijke taak bestond uit het onthouden van drie getallen, de moeilijke taak bestond uit het onthouden van zes getallen. Na het lezen van het verhaaltje moesten deze getallen worden herhaald. Wanneer de tweede taak makkelijk was, dan hadden de deelnemers meer werkgeheugencapaciteit over om het verhaaltje te begrijpen. Wanneer daarentegen de tweede taak moeilijk was, dan hadden deelnemers minder werkgeheugencapaciteit over om het verhaaltje te begrijpen. De laatste zin van de verhaaltjes begon met het potentieel ambigue voornaamwoord *hij* of *zij*. De resultaten van het experiment laten zien dat bij het uitvoeren van een *moeilijke* tweede taak volwassen deelnemers vaker voor een niet-prominente referent kozen dan bij het uitvoeren van een *makkelijke* tweede taak. De reden hiervoor is dat ze bij een moeilijke tweede taak minder gebruik maakten van de informatie uit de vorige zin.

Met andere woorden, met een hogere belasting van het werkgeheugen lijken volwassenen meer op kinderen in het verwerken van voornaamwoorden. Dit ondersteunt de voorspelling van het model dat voor het verwerken van voornaamwoorden voldoende werkgeheugencapaciteit nodig is.

Psychofysiologische experimenten

In aanvulling op de twee hierboven beschreven gedragsexperimenten hebben we twee psychofysiologische experimenten uitgevoerd om te onderzoeken wat er gebeurt tijdens het verwerken van voornaamwoorden.

Psychofysiologische experimenten. In een psychofysiologisch experiment wordt getest of verschillende condities, bijvoorbeeld het interpreteren van zinnen in verschillende contexten, ook leiden tot verschillen in fysiologische signalen, zoals het registreren van elektrische activiteit in de hersenen (EEG) of pupilgrootte. In het geval dat er verschillen worden gevonden in het fysiologische signaal als gevolg van de verschillende condities, dan is de volgende vraag op welk tijdstip de signalen van elkaar beginnen te verschillen. Psychofysiologische metingen kunnen op deze manier een preciezer beeld geven van de processen die optreden tijdens het verwerken van een zin.

In deze sectie worden een EEG-experiment en een pupilgrootte-experiment beschreven waarmee twee aannames van de computermodellen zijn getest.

Experiment 3: Effect van context op het begrijpen van voornaamwoorden in objectpositie

In de zinnen *Gisteren sprak James met Rob. Hij bekende de diefstal.* is het persoonlijke voornaamwoord *hij* ambigu, omdat er twee mogelijke referenten in de context aanwezig zijn: James en Rob. De vraag is of volwassen deelnemers bij het lezen van een voornaamwoord eerst terug

gaan naar de talige context om een keuze te maken tussen verschillende mogelijke referenten, of dat ze al meteen bij het lezen van een voornaamwoord de meest plausibele referent paraat hebben.

Hoofdstuk 5 presenteert een experiment waarin we deze vraag hebben onderzocht met een vergelijkbare dubbele taak als het vorige experiment: aan volwassen deelnemers werd gevraagd om korte verhaaltjes te lezen, terwijl ze tegelijkertijd een tweede taak moesten uitvoeren. Deze tweede taak kon moeilijk of makkelijk zijn. De laatste zin van de verhaaltjes begon met een mogelijk ambigu voornaamwoord *hij* of *zij*. In dit experiment werden via elektrodes op de hoofdhuid van de deelnemer elektrische signalen geregistreerd tijdens het uitvoeren van de dubbele taak. We hebben onderzocht of de elektrische activiteit die ontstaat in de hersenen tijdens het lezen van de verhaaltjes verschillend zijn afhankelijk van de zinsstructuur van de context en de moeilijkheid van de tweede taak.

De resultaten van dit experiment laten zien dat deelnemers bij het lezen van een ambigu voornaamwoord niet alle mogelijke referenten overwegen, maar vaak al meteen weten wie er wordt bedoeld. Voor de deelnemers was al snel tijdens het lezen van een verhaaltje duidelijk wie de meest prominente persoon was in dit verhaaltje. In het begin van het lezen van het verhaaltje werd er een effect gemeten van de moeilijkheid van de tweede taak: deelnemers lijken de informatie uit de voorgaande zinnen meer te gebruiken bij een makkelijke tweede taak dan bij een moeilijke tweede taak. Maar aan het eind van het verhaaltje, tijdens het lezen van het voornaamwoord, waren deze verschillen in EEG signalen bijna verdwenen.

Experiment 4: Effect van context op het begrijpen van voornaamwoorden in subjectpositie

In onze computermodellen wordt hetzelfde onderliggende mechanisme gebruikt voor de verwerking van voornaamwoorden in subjectpositie (*hij*, *zij*) als voor voornaamwoorden in objectpositie (*hem*, *haar*). Omdat de context een vroeg effect blijkt te hebben op de verwerking van voornaamwoorden in subjectpositie, volgt de voorspelling dat de context ook een vroeg effect kan hebben op de verwerking van voornaamwoorden in objectpositie. In talen zoals het Nederlands en Engels wordt de interpretatie van voornaamwoorden in objectpositie sterk bepaald door grammaticale principes op zinsniveau. Daarom zal de context niet snel de uiteindelijke interpretatie van voornaamwoorden beïnvloeden, maar mogelijk wel het proces waardoor die interpretatie gevonden wordt.

Hoofdstuk 6 beschrijft een experiment om het effect van de context op de verwerking van voornaamwoorden in objectpositie te onderzoeken, waarbij de pupilgrootte wordt gemeten. Het groter of kleiner worden van de pupil is een onbewuste reflex en treedt bijvoorbeeld op bij fel licht, met als doel de lichtinval te blokkeren. Maar ook bij een constante lichtbron kunnen kleine wijzigingen in pupilgrootte worden gemeten ten gevolge van cognitieve activiteit, geheugenbelasting of emotionele reacties.

In het vierde experiment binnen dit onderzoek hebben we de pupilgrootte gemeten tijdens de verwerking van zinnen als *Hier zie je een pinguïn en een schaap. De pinguïn slaat hem met een pan.* We hebben onderzocht of de verwerking van het voornaamwoord *hem* moeilijker of

makkelijker wordt als de referenten in omgekeerde volgorde worden geïntroduceerd: *Hier zie je een schaap en een pinguïn....* Als de context een effect heeft op de interpretatie van voornaamwoorden, dan verwachtten we dat eerstgenoemde referenten, die prominenter zijn, eerder overwogen worden dan later genoemde referenten. Tijdens het horen van de zin zagen de deelnemers een afbeelding die de correcte interpretatie van de zin weergaf (een pinguïn die een schaap sloeg) of een afbeelding die een incorrecte interpretatie van de zin weergaf (een pinguïn die zichzelf sloeg).

De resultaten van het experiment laten zien dat tijdens het horen van het voornaamwoord zowel de volgorde van introductie van de referenten als de passendheid van de zin bij het plaatje effect hebben op de pupilgrootte. Wanneer de introductie zin begint met de pinguïn, dan overwegen de deelnemers in eerste instantie de interpretatie die ze zien op het plaatje. Dit leidt tot een lagere pupilgrootte (minder cognitieve activiteit) als het plaatje correct is, maar tot een grotere pupilgrootte (meer cognitieve activiteit) als het plaatje incorrect is. Wanneer de introductie zin begint met het schaap, dan is er geen verschil in pupilgrootte. Waarschijnlijk is er hier geen verschil, omdat deelnemers in dit geval niet vertrouwen op het plaatje en dus liever wachten met het maken van een keus tot ze de rest van de zin hebben gehoord.

De verschillen in pupilgrootte tussen de condities ontstaan al erg vroeg tijdens de verwerking van het voornaamwoord. Dit ondersteunt de voorspelling van het model dat context een vergelijkbare rol speelt in de verwerking van voornaamwoorden in objectpositie als in de verwerking van voornaamwoorden in subjectpositie.

Conclusies

In dit onderzoek is onderzocht waarom kinderen er zo lang over doen om het gebruik en de interpretatie van persoonlijke voornaamwoorden onder de knie te krijgen, en welke algemene cognitieve factoren van invloed zijn op de volwassen verwerking van voornaamwoorden.

De resultaten van de twee cognitieve modellen en de vier experimentele studies die zijn gepresenteerd in dit proefschrift wijzen erop dat de interpretatie en het begrip van voornaamwoorden voldoende verwerkingssnelheid en voldoende werkgeheugencapaciteit vereisen. Door onvoldoende verwerkingssnelheid en onvoldoende werkgeheugencapaciteit maken Nederlandstalige en Engelstalige kinderen nog tot zeker hun zesde jaar fouten met verschillenden soorten voornaamwoorden. Maar ook volwassen kunnen vergelijkbare fouten maken wanneer hun werkgeheugen zwaar belast is.

Een ander interessant resultaat van dit onderzoek is dat de context een vroeg effect heeft op het begrijpen van voornaamwoorden. Volwassenen hebben meestal geen problemen met de ambiguïteit van voornaamwoorden, omdat de structuur van de talige context de luisteraar vaak voldoende aanwijzingen geeft om de bedoelde referent te achterhalen. Kortom, volwassenen gebruiken de talige context om voorspellingen te doen over hoe de zin of het verhaal verder zal gaan.

Hoofdstuk 7 bespreekt de verschillende implicaties van dit onderzoek. Een belangrijk onderwerp in normale taalontwikkeling is hoe kinderen leren waar persoonlijke voornaam-

woorden naar verwijzen en hoe deze verwijzende woorden worden gebruikt. De resultaten geven meer inzicht in dit onderwerp. Daarnaast geven de resultaten meer inzicht in de mechanismes die volwassen voor het verwerken van taal, en in de rol van context in zinsverwerking. Het is belangrijk om meer te weten over deze mechanismes, omdat moeilijkheden met het verwerken van voornaamwoorden door bijvoorbeeld cognitieve problemen of een taalstoornis kunnen leiden tot problemen in de communicatie.

Groningen Dissertations in Linguistics

1. Henriëtte de Swart (1991). *Adverbs of Quantification: A Generalized Quantifier Approach.*
2. Eric Hoekstra (1991). *Licensing Conditions on Phrase Structure.*
3. Dicky Gilbers (1992). *Phonological Networks. A Theory of Segment Representation.*
4. Helen de Hoop (1992). *Case Configuration and Noun Phrase Interpretation.*
5. Gosse Bouma (1993). *Nonmonotonicity and Categorial Unification Grammar.*
6. Peter I. Blok (1993). *The Interpretation of Focus.*
7. Roelien Bastiaanse (1993). *Studies in Aphasia.*
8. Bert Bos (1993). *Rapid User Interface Development with the Script Language Gist.*
9. Wim Kosmeijer (1993). *Barriers and Licensing.*
10. Jan-Wouter Zwart (1993). *Dutch Syntax: A Minimalist Approach.*
11. Mark Kas (1993). *Essays on Boolean Functions and Negative Polarity.*
12. Ton van der Wouden (1994). *Negative Contexts.*
13. Joop Houtman (1994). *Coordination and Constituency: A Study in Categorial Grammar.*
14. Petra Hendriks (1995). *Comparatives and Categorial Grammar.*
15. Maarten de Wind (1995). *Inversion in French.*
16. Jelly Julia de Jong (1996). *The Case of Bound Pronouns in Peripheral Romance.*
17. Sjoukje van der Wal (1996). *Negative Polarity Items and Negation: Tandem Acquisition.*
18. Anastasia Giannakidou (1997). *The Landscape of Polarity Items.*
19. Karen Lattewitz (1997). *Adjacency in Dutch and German.*
20. Edith Kaan (1997). *Processing Subject-Object Ambiguities in Dutch.*
21. Henny Klein (1997). *Adverbs of Degree in Dutch.*
22. Leonie Bosveld-de Smet (1998). *On Mass and Plural Quantification: The case of French 'des'/'du'-NPs.*
23. Rita Landeweerd (1998). *Discourse semantics of perspective and temporal structure.*
24. Mettina Veenstra (1998). *Formalizing the Minimalist Program.*
25. Roel Jonkers (1998). *Comprehension and Production of Verbs in aphasic Speakers.*
26. Erik F. Tjong Kim Sang (1998). *Machine Learning of Phonotactics.*
27. Paulien Rijkhoek (1998). *On Degree Phrases and Result Clauses.*
28. Jan de Jong (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure.*
29. H. Wee (1999). *Definite Focus.*

30. Eun-Hee Lee (2000). *Dynamic and Stative Information in Temporal Reasoning: Korean tense and aspect in discourse.*
31. Ivilin P. Stoianov (2001). *Connectionist Lexical Processing.*
32. Klarien van der Linde (2001). *Sonority substitutions.*
33. Monique Lamers (2001). *Sentence processing: using syntactic, semantic, and thematic information.*
34. Shalom Zuckerman (2001). *The Acquisition of "Optional" Movement.*
35. Rob Koeling (2001). *Dialogue-Based Disambiguation: Using Dialogue Status to Improve Speech Understanding.*
36. Esther Ruigendijk (2002). *Case assignment in Agrammatism: a cross-linguistic study.*
37. Tony Mullen (2002). *An Investigation into Compositional Features and Feature Merging for Maximum Entropy-Based Parse Selection.*
38. Nanette Bienfait (2002). *Grammatica-onderwijs aan allochtone jongeren.*
39. Dirk-Bart den Ouden (2002). *Phonology in Aphasia: Syllables and segments in level-specific deficits.*
40. Rienk Withaar (2002). *The Role of the Phonological Loop in Sentence Comprehension.*
41. Kim Sauter (2002). *Transfer and Access to Universal Grammar in Adult Second Language Acquisition.*
42. Laura Sabourin (2003). *Grammatical Gender and Second Language Processing: An ERP Study.*
43. Hein van Schie (2003). *Visual Semantics.*
44. Lilia Schürcks-Grozeva (2003). *Binding and Bulgarian.*
45. Stasinos Konstantopoulos (2003). *Using ILP to Learn Local Linguistic Structures.*
46. Wilbert Heeringa (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance.*
47. Wouter Jansen (2004). *Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology.*
48. Judith Rispens (2004). *Syntactic and phonological processing in developmental dyslexia.*
49. Danielle Bougaïré (2004). *L'approche communicative des campagnes de sensibilisation en santé publique au Burkina Faso: Les cas de la planification familiale, du sida et de l'excision.*
50. Tanja Gaustad (2004). *Linguistic Knowledge and Word Sense Disambiguation.*
51. Susanne Schoof (2004). *An HPSG Account of Nonfinite Verbal Complements in Latin.*
52. M. Begoña Villada Moirón (2005). *Data-driven identification of fixed expressions and their modifiability.*
53. Robbert Prins (2005). *Finite-State Pre-Processing for Natural Language Analysis.*
54. Leonoor van der Beek (2005). *Topics in Corpus-Based Dutch Syntax*
55. Keiko Yoshioka (2005). *Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse.*
56. Sible Andringa (2005). *Form-focused instruction and the development of second language proficiency.*
57. Joanneke Prenger (2005). *Taal telt! Een onderzoek naar de rol van taalvaardigheid en tekstbegrip in het realistisch wiskundeonderwijs.*
58. Neslihan Kansu-Yetkiner (2006). *Blood, Shame and Fear: Self-Presentation Strategies of Turkish Women's Talk about their Health and Sexuality.*
59. Mónika Z. Zempléni (2006). *Functional imaging of the hemispheric contribution to language processing.*
60. Maartje Schreuder (2006). *Prosodic Processes in Language and Music.*
61. Hidetoshi Shiraishi (2006). *Topics in Nivkh Phonology.*
62. Tamás Biró (2006). *Finding the Right Words: Implementing Optimality Theory with Simulated Annealing.*

63. Dieuwke de Goede (2006). *Verbs in Spoken Sentence Processing: Unraveling the Activation Pattern of the Matrix Verb.*
64. Eleonora Rossi (2007). *Clitic production in Italian agrammatism.*
65. Holger Hopp (2007). *Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing.*
66. Gerlof Bouma (2008). *Starting a Sentence in Dutch: A corpus study of subject- and object-fronting.*
67. Julia Klitsch (2008). *Open your eyes and listen carefully. Auditory and audiovisual speech perception and the McGurk effect in Dutch speakers with and without aphasia.*
68. Janneke ter Beek (2008). *Restructuring and Infinitival Complements in Dutch.*
69. Jori Mur (2008). *Off-line Answer Extraction for Question Answering.*
70. Lonneke van der Plas (2008). *Automatic Lexico-Semantic Acquisition for Question Answering.*
71. Arjen Versloot (2008). *Mechanisms of Language Change: Vowel reduction in 15th century West Frisian.*
72. Ismail Fahmi (2009). *Automatic term and Relation Extraction for Medical Question Answering System.*
73. Tuba Yarbay Duman (2009). *Turkish Agrammatic Aphasia: Word Order, Time Reference and Case.*
74. Maria Trofimova (2009). *Case Assignment by Prepositions in Russian Aphasia.*
75. Rasmus Steinkrauss (2009). *Frequency and Function in WH Question Acquisition. A Usage-Based Case Study of German L1 Acquisition.*
76. Marjolein Deunk (2009). *Discourse Practices in Preschool. Young Children's Participation in Everyday Classroom Activities.*
77. Sake Jager (2009). *Towards ICT-Integrated Language Learning: Developing an Implementation Framework in terms of Pedagogy, Technology and Environment.*
78. Francisco Dellatorre Borges (2010). *Parse Selection with Support Vector Machines.*
79. Geoffrey Andogah (2010). *Geographically Constrained Information Retrieval.*
80. Jacqueline van Kruiningen (2010). *Onderwijsontwerp als conversatie. Probleemoplossing in interprofessioneel overleg.*
81. Robert G. Shackleton (2010). *Quantitative Assessment of English-American Speech Relationships.*
82. Tim Van de Cruys (2010). *Mining for Meaning: The Extraction of Lexico-semantic Knowledge from Text.*
83. Therese Leinonen (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects.*
84. Erik-Jan Smits (2010). *Acquiring Quantification. How Children Use Semantics and Pragmatics to Constrain Meaning.*
85. Tal Caspi (2010). *A Dynamic Perspective on Second Language Development.*
86. Teodora Mehotcheva (2010). *After the fiesta is over. Foreign language attrition of Spanish in Dutch and German Erasmus Student.*
87. Xiaoyan Xu (2010). *English language attrition and retention in Chinese and Dutch university students.*
88. Jelena Prokić (2010). *Families and Resemblances.*
89. Radek Šimík (2011). *Modal existential wh-constructions.*
90. Katrien Colman (2011). *Behavioral and neuroimaging studies on language processing in Dutch speakers with Parkinson's disease.*
91. Siti Mina Tamah (2011). *A Study on Student Interaction in the Implementation of the Jigsaw Technique in Language Teaching.*

92. Aletta Kwant (2011). *Geraakt door prentenboeken. Effecten van het gebruik van prentenboeken op de sociaal-emotionele ontwikkeling van kleuters.*
93. Marlies Kluck (2011). *Sentence amalgamation.*
94. Anja Schüppert (2011). *Origin of asymmetry: Mutual intelligibility of spoken Danish and Swedish.*
95. Peter Nabende (2011). *Applying Dynamic Bayesian Networks in Transliteration Detection and Generation.*
96. Barbara Plank (2011). *Domain Adaptation for Parsing.*
97. Çağri Çöltekin (2011). *Catching Words in a Stream of Speech: Computational simulations of segmenting transcribed child-directed speech.*
98. Dörte Hessler (2011). *Audiovisual Processing in Aphasic and Non-Brain-Damaged Listeners: The Whole is More than the Sum of its Parts.*
99. Herman Heringa (2012). *Appositional constructions.*
100. Diana Dimitrova (2012). *Neural Correlates of Prosody and Information Structure.*
101. Harwintha Anjarningsih (2012). *Time Reference in Standard Indonesian Agrammatic Aphasia.*
102. Myrte Gosen (2012). *Tracing learning in interaction. An analysis of shared reading of picture books at kindergarten.*
103. Martijn Wieling (2012). *A Quantitative Approach to Social and Geographical Dialect Variation.*
104. Gisi Cannizzaro (2012). *Early word order and animacy.*
105. Kostadin Cholakov (2012). *Lexical Acquisition for Computational Grammars. A Unified Model.*
106. Karin Beijering (2012). *Expressions of epistemic modality in Mainland Scandinavian. A study into the lexicalization-grammaticalization-pragmaticalization interface.*
107. Veerle Baaijen (2012). *The development of understanding through writing.*
108. Jacolien van Rij (2012). *Pronoun processing: Computational, behavioral, and psychophysiological studies in children and adults.*

GRODIL
Center for Language and Cognition Groningen (CLCG)
P.O. Box 716
9700 AS Groningen
The Netherlands