

ABSTRACT

Title of Document: **MEMORY AND PREDICTION IN CROSS-LINGUISTIC SENTENCE COMPREHENSION**

Maria S. Lago
Doctor of Philosophy, 2014

Directed By: Professor Colin Phillips
Department of Linguistics

This dissertation explores the role of morphological and syntactic variation in sentence comprehension across languages. While most previous research has focused on how cross-linguistic differences affect the control structure of the language architecture (Lewis & Vasishth, 2005) here we adopt an explicit model of memory, content-addressable memory (Lewis & Vasishth, 2005; McElree, 2006) and examine how cross-linguistic variation affects the nature of the representations and processes that speakers deploy during comprehension. With this goal, we focus on two kinds of grammatical dependencies that involve an interaction between language and memory: subject-verb agreement and referential pronouns. In the first part of this dissertation, we use the self-paced reading method to examine how the processing of subject-verb agreement in Spanish, a language with a rich morphological system, differs from English. We show that differences in morphological richness across languages impact prediction processes while leaving retrieval processes fairly preserved. In the second

part, we examine the processing of coreference in German, a language that, in contrast with English, encodes gender syntactically. We use eye-tracking to compare comprehension profiles during coreference and we find that only speakers of German show evidence of semantic reactivation of a pronoun's antecedent. This suggests that retrieval of semantic information is dependent on syntactic gender, and demonstrates that German and English speakers retrieve qualitatively different antecedent representations from memory. Taken together, these results suggest that cross-linguistic variation in comprehension is more affected by the content than the functional importance of gender and number features across languages.

MEMORY AND PREDICTION IN CROSS-LINGUISTIC SENTENCE
COMPREHENSION

By

Maria Sol Lago

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Colin Phillips, Chair
Professor Ellen Lau
Professor Alexander Williams
Professor Yi-Ting Huang
Professor Yasmeen Faroqi Shah

© Copyright by
Maria Sol Lago
2014

Dedication

To my parents

Acknowledgements

These five years have been the most challenging and nerve-racking years of my life so far. And they have been truly great. And while I am sure that I started this process, I also know that I would not have gotten to this point without the support of an entire community, both at Maryland and in Buenos Aires. This is my attempt to acknowledge some of the members of this community, but the list is far from exhaustive. And in the end, I wish that I could just get all of them in a room and thank them for inspiring me, for teaching me linguistics, and for allowing me to carry out the research that I love.

First and foremost, there is Colin Phillips. Colin has been my harshest critic and my most firm supporter. From him I learned to always aim higher, and to carefully think about why what I do should matter to other people. I still have no clue of why he took me as his student, but I am very grateful that he did. I will say here something that I have read in many dissertations before: I really hope that one day I can become as good a mentor to my students as Colin has been to me.

Ellen Lau is not only my co-advisor but also the most enthusiastic researcher I know. After innumerable meetings, she has become the voice in my head that forces me to carefully walk through all possible outcomes of an experiment and to think carefully of the hypotheses at stake. From Ellen I take a constant passion for trying out different techniques and analyses, and the ability to use findings from multiple domains to enrich the way in which we think about language.

Together with Colin and Ellen, I would like to thank Yasmeen Faroqi Shah, Yi-Ting Huang and Alexander Williams for serving in my dissertation committee. And I would like to thank the faculty in the Department of Linguistics, who to my eyes has always looked like the Justice League of linguistics. Now that I am about to leave, I realize that trying to reproduce the many ways in which they are unique will probably take me a lifetime. But just in case, I will put down these features as a reminder to myself: I hope that one day I will have Norbert Hornstein's ability to make interesting points in any discussion, Jeff Lidz's careful way of walking through an argument, Bill Idsardi's endless knowledge about phonology and cognitive science, Alexander William's curiosity and enthusiasm for detail and Naomi Feldman's honest and systematic way of conveying ideas. And of course, I also hope to become as good a teacher as Howard Lasnik has been to me.

Wing Yee Chow has taught me most of what I know about experimental design. And while doing so, she has become an indispensable part of my life outside grad school. The thing with Wing Yee is that working and playing are one and the same thing with her, and my enthusiasm for research is strongly rooted in the endless debates, discussions and brainstorming sessions that we have had over the years. Together with her, Ewan Dunbar was an integral part of my life at Maryland. Ewan taught me

how to script, and most importantly, he taught me that numbers and math should make sense. From him I take an often times painful passion for clarity in thinking, and the belief that the way in which we analyze data should convey the way in which we think about them.

My cohort has also been a constant source of inspiration and support. Dan Parker, my officemate, made coming to work a pleasant experience, and I owe him for many discussions at the whiteboard, lessons about rap and later, when writing times arrived, quite a lot of lemon ginger cookies. Yakov Kronrod puzzled over phonology with me, and then took me out on motor-rides and road trips. I am very happy to have shared my life in College Park with him. I also want to thank the other people in my year, who have been great classmates and supportive coworkers: Angela Xiaoxue He, Kenshi Funakoshi, Darryl McAdams, Alexis Wellwood and Megan Sutton. Of the students in the year before me, I am glad to have shared many evenings with Shevaun Lewis, who is a formidable discussion partner and an amazing cook, and Dave Kush, whose fascinating personality resists any summarizing attempts. My first year at Maryland would not have been the same without the presence of Dave and Shannon Barrios, who was a kind and encouraging roommate. And my last years here have definitely benefitted from lively interactions with Shota Momma, Dustin Chacón, Zoe Schlueter, Tim Dawson, and Anna Bonnet.

Outside Maryland, there are several people who I am lucky to simultaneously consider friends and collaborators. Mathias Scharinger taught me about magnetoencephalography and dipoles, and he also provided endless amounts of chocolate, companionship and fun trips during these years. I honestly don't know how I would have survived in the US without Mathias, and I am looking forward to many more years of research together. Silvia Marijuan has been a good friend since college. She has taught me about bilingualism and, together with Scott Ferree, provided me with healthy amounts of meat and lively discussion, so I would not forget my Argentinian origins. I was lucky to share a semester with Carmen Gallar Sánchez, who can make up a syntactic theory while cooking dinner and then argue about it until dessert time. During my last semester, I met Anamaria Bentea, who made sure that I went to the gym regularly and took me out to enjoy good food and her exciting company when I felt overwhelmed by writing. I also would like to thank Morgan Moyer, who is an adventurous friend, and Naho Orita, with whom I enjoyed many cheerful dinners when we lived together during my last year at Maryland.

Of the methods that I acquired at Maryland, I am particularly grateful to have gotten to learn eye-tracking. I owe a lot of my passion for this technique to my interactions with a group of talented and thoughtful researchers who I strongly admire. Among them, I want to thank Akira Omaki and Brian Dillon, together with Alan Mishler, Shayne Sloggett and Erin Mahoney, who have worked with me over the years. I also want to thank Ilia Kurenkov, who not only helped me figure out the settings of our eye-tracker, but who was also a relentless rowing partner down the Potomac River. From outside my lab, I am grateful to have enjoyed stimulating discussions with Peter

Gordon, Adrian Staub, Mariano Sigman, Diego Shalom, Lena Jäger and multiple people from Shravan Vasishth's group at Potsdam.

Finally, there are all the people back at home in Buenos Aires. They are the people that I grew up with, but also the people who, every summer, remind me who I am and keep me grounded: Anabel Miguelez, Juan Pablo Pekarek, Lucía Levis, Ana Zelzman, Mariela Penna, Omar Yabur, Eleonora Pinotti and Carolina Gattei. And finally, to my family and especially to my parents, what can I say? You made me who I am, and you are responsible for teaching me that it is important to take a stand about things, and to never ever give up trying.

Table of Contents

Dedication	ii
Acknowledgements	iii
List of Tables.....	viii
List of Figures	ix
1 Introduction.....	1
1.1 Representations and mechanisms cross-linguistically.....	1
1.2 A cognitive architecture for retrieval	5
1.3 Two test cases of cross-linguistic variation.....	9
1.3.1 Agreement	9
1.3.2 Coreference	12
1.4 Outline of the dissertation.....	14
2 Agreement processing in a morphologically rich language	17
2.1 Outline.....	17
2.2 Introduction.....	18
2.3 Experiment 1: Spanish Main Verbs	29
2.3.1 Methods.....	29
2.3.2 Results.....	37
2.3.3 Discussion.....	41
2.4 Experiment 2: Spanish Auxiliaries.....	44
2.4.1 Methods.....	44
2.4.2 Results.....	45
2.4.3 Discussion.....	48
2.5 Experiment 3: English Auxiliaries.....	52
2.5.1 Methods.....	53
2.5.2 Results.....	54
2.5.3 Discussion.....	57
2.6 General Discussion.....	58
2.7 Conclusion	70
3 Coreference and the lack of antecedent frequency effects in English... 	71
3.1 Outline.....	71
3.2 Antecedent representation in English.....	72
3.3 Antecedent representation in languages with syntactic gender	79
3.4 van Gompel & Majid (2004).....	83
3.5 Experiment 4: No frequency effects with common noun antecedents	87
3.5.1 Methods.....	89
3.5.2 Results.....	95
3.5.3 Discussion.....	100
3.6 Experiment 5: No frequency effects with proper name antecedents	101
3.6.1 Methods.....	101
3.6.2. Results.....	104
3.6.3. Discussion.....	106
3.7 Conclusion	107
4 Antecedent representation in a language with syntactic gender	115
4.1. Outline.....	115

4.2 Introduction.....	116
4.3 Experiment 6: Antecedent representation in German.....	126
4.3.1 Methods	126
4.3.2 Results.....	132
4.3.3 Discussion.....	139
4.4 Experiment 7: Antecedent representation in English	143
4.4.1 Methods.....	146
4.4.2 Results.....	150
4.4.3 Discussion.....	157
4.5 General Discussion.....	158
4.6 Conclusion	165
5 Conclusions	166
5.1 Specific conclusions.....	166
5.1.1 Agreement.....	166
5.1.2 Coreference.....	171
5.2 Overall conclusions.....	177
Appendices	181
Bibliography	211

List of Tables

Table 1 Sample set of experimental items in Experiments 1 and 2	31
Table 2 Linear mixed-effects model results for the regions of interest in Experiment 1. Significant effects are bolded.....	39
Table 3 Linear mixed-effects model results for the regions of interest in Experiment 2. Significant effects are bolded.....	47
Table 4 Sample set of experimental items in Experiment 3.....	54
Table 5 Linear mixed-effects model results for the regions of interest in Experiment 3. Significant effects are bolded.....	56
Table 6 Sample Materials and delimited regions from Experiment 4.....	91
Table 7 Lexical frequency of antecedents (per million words).....	92
Table 8 Averages in first fixation duration, first-pass duration, and total time across critical regions in Experiment 4.....	96
Table 9 Statistics for first fixation, first-pass and total time measures in Experiment 4. Reliable effects are in bold font, marginal effects with $.05 < p < .07$ are in bold and italic font.	96
Table 10 Fitted ex-Gaussian parameter values in Experiment 4, along with the sample means for comparison. Reliable differences are in bold.	97
Table 11 A sample item in Experiment 5 with critical regions indicated.	103
Table 12 Averages in first fixation duration, first-pass duration, and total time across critical regions in Experiment 5.....	105
Table 13 Statistics for first fixation, first-pass and total time measures in Experiment 5. Reliable effects are in bold font, marginal effects with $.05 < p < .07$ are in bold and italic font.	105
Table 14 Sample set of experimental items and delimited regions in the German experiment.....	128
Table 15 Region averages and standard errors in milliseconds in the German experiment.....	133
Table 16 Statistics for reading time measures in the German experiment. Reliable effects are in bold font.	134
Table 17 Region averages and standard errors in milliseconds in the English experiment.....	151
Table 18 Statistics for reading time measures in the English experiment. Reliable effects are in bold font.	152

List of Figures

Figure 1 Results in Sachs (1967). <i>Left panel</i> : percent correct responses on the “identical vs. changed” judgment task. <i>Right panel</i> : participants’ self-reported confidence in their judgments. Reproduced from Wagers (2007).....	8
Figure 2 Use of root infinitives across languages. Children are separated into 3 groups according to the richness of the inflectional paradigm in their language. Black markers represent English and Swedish, which distinguish no more than two forms in any tense. Proportions of root infinitives are highest among children in this group. Unfilled markers represent French, German, and Dutch, which distinguish up to 3 or 4 forms for regular main verbs in any tense. Proportions of root infinitives are markedly lower for this group. Finally, grey markers represent Spanish, Catalan, Hebrew, and Italian, which show the richest agreement paradigms. Proportions of root infinitives are lowest in this group. Reproduced from Phillips (2010).....	11
Figure 3 Region by region means in Experiment 1. Error bars indicate the standard error of the mean. Sample sentence: <i>The note(s) that the girl wrote_{sg/pl} during class cheered her friend up</i> . The regions of interest are underlined.....	38
Figure 4 Vincentile plots for the four experimental conditions in the verb+1 region in Experiment 1. Error bars indicate the standard error of the mean across participants.....	40
Figure 5 Region by region means in Experiment 2. Error bars indicate the standard error of the mean. Sample sentence: <i>The note(s) that the girl is/are going to write during class will cheer her friend up</i> . The regions of interest are underlined.....	46
Figure 6 Vincentile plots for the four experimental conditions in the verb+1 region in Experiment 2. Error bars indicate the standard error of the mean across participants.....	48
Figure 7 Region by region means in Experiment 3. Error bars indicate the standard error of the mean. The regions of interest are underlined.....	55
Figure 8 Vincentile plots for the four experimental conditions in verb+1 region in Experiment 3. Error bars indicate the standard error of the mean across participants.....	57
Figure 9 Comparison of grammaticality and attraction effect sizes across four experiments using a standardized scale (Cohen’s d). The effect of grammaticality was calculated as the mean RT difference between grammatical and ungrammatical sentences in the no attractor conditions. The effect of attraction was calculated as the mean RT difference between the singular and plural noun conditions in ungrammatical sentences. Diamonds show mean effect sizes, and vertical bars display bootstrapped 95% CIs. Effect sizes were always estimated from length-regressed RTs in the post-verbal region. Data for English main verbs was obtained from WL&P (2009: Experiment 2)	61

Figure 10 Predicted ex-Gaussian density plots based on parameter values averaged across participants from Experiment 4. The high frequency condition is plotted as a solid line, and the low frequency condition is plotted as a dashed line. The regions of interest are the antecedent (<i>captain/mariner</i>), anaphor (<i>ut his</i>) and post-anaphor (<i>safety</i>) regions.....	100
Figure 11 A cognitive architecture for the lexicon. Levelt's model proposes a feed forward activation-spreading network with three levels. Nodes in the top conceptual level represent lexical concepts. Nodes in the lemma level represent syntactic properties of words. Nodes in the lexeme level represent phonological and phonetic properties. Left panel: processing stages in the network. Right panel: example fragment of lexical network. Lemmas are shaded in grey and gender information corresponds to the Dutch equivalents of the English words. Adapted from Levelt (1998: 171).....	110
Figure 12 Possible scenarios in antecedent retrieval across languages. LSG stands for languages with syntactic gender.....	114
Figure 13 Semantic association effects in the German experiment. Mean difference scores and their standard error are shown with squares and bars respectively. Difference scores were computed as the mean difference between the related and the unrelated conditions. Therefore, negative difference scores index shorter reading times in the related than unrelated conditions (a priming effect).	137
Figure 14 Reading time profiles in the English experiment. Mean difference scores and their standard error are shown with squares and bars respectively. Difference scores were computed as the mean difference between the related and the unrelated conditions. Therefore, negative difference scores index shorter reading times in the related than unrelated conditions (a priming effect).	155

1 Introduction

1.1 Representations and mechanisms cross-linguistically

The study of cross-linguistic variation in psycholinguistics is marked by a tension. On the one hand, any theory of language processing aims to describe the properties of the cognitive architecture that allows speakers to produce and understand language. Many of these properties should be shared across languages if this architecture is innate and specific to humans. On the other hand, it is known that speakers adapt to their native language. Thus, it is likely that their cognitive architecture will be optimized to handle the phonological, morphological, syntactic and semantic properties of the language that they acquire from infancy. As a result, studying processing across languages involves trying to achieve a balance between the theoretical goal of describing a generalizable architecture and the acknowledgment that if this architecture is to be adaptive, then some variation needs to occur.

In this thesis I adopt the idea that a computationally complete model of the language architecture should jointly describe the *processes*, *representations*, and *control structures* used in language processing (Lewis & Vasishth, 2005). Representations are the temporary pieces of information that are relevant for a given parsing task, processes refer to operations over that information, and control structures are the decision principles that determine when different processes are applied. My work applies this framework to the study of cross-linguistic variation and addresses the following question: *how do systematic changes in grammatical features*

across languages affect the representations and processes involved in sentence comprehension?

From this point of view, previous approaches to cross-linguistic variation can be understood as characterizing the control structure of the language architecture. For example, most of these approaches have tried to account for the principles that guide speakers' attachment decisions in sentences such as the one in (1):

- (1) Someone shot the servant of the actress who was on the balcony.

This sentence is ambiguous because the relative clause “who was on the balcony” can modify either the syntactically higher noun phrase “the servant”, or the lower noun phrase “the actress”. Furthermore, there is variation in how speakers of different languages resolve this ambiguity: English speakers show a reliable preference for attaching the RC low (i.e., in questionnaire studies when they are asked who was on the balcony they select “actress” on 58% of trials), whereas Spanish speakers opt predominantly for the higher attachment site, by choosing “servant” on 62% of trials (Cuetos & Mitchell, 1988; see also Carreiras, 1992; Carreiras & Clifton, 1993; Gilboy, Sopena, Clifton, & Frazier, 1995; Mitchell & Cuetos, 1991; Mitchell, Cuetos, & Zagar, 1990).

Previous accounts of this variability fall into three major classes. Some accounts, sometimes called “universalist approaches”, propose that attachment decisions rely only on syntactic principles. These principles serve to increase the speed and efficiency with which new material can be incorporated into the analysis of

a sentence (e.g., Bever, 1970; Crocker, 1992; Frazier, 1987; Inoue & Fodor, 1995; Kimball, 1973). They are claimed to apply universally across languages, as they should produce benefits for all types of readers and listeners. Since universalist approaches claim that speakers' attachment decisions are only guided by uniform syntactic principles, they attribute cross-linguistic variation in processing to the effects of later reanalysis processes that take into account pragmatic and thematic factors.

On the other hand, a second class of accounts, usually referred to as "parameterized models", argues that the principles that govern attachment preferences can be language-dependent (e.g., Gibson, Pearlmuter, Canseco-Gonzales, & Hickock, 1996; Gibson, & Schütze, 1999; Hemforth, Konieczny, Scheepers, & Strube, 1988; Mazuka and Lust, 1990). They propose that these principles are represented through a small number of parameters that are set differently across languages during acquisition. Therefore, in contrast with universalist approaches, they allow for cross-linguistic variation at the processes-level but propose that this variation is constrained in systematic ways in each language. Finally, a third type of accounts, usually identified as statistical or experience-based accounts, propose that attachment preferences are determined by speakers' exposure to comparable examples in their prior experience with the language. These accounts assign a key role to the statistical frequency of structures in a language, which are thought to determine how reliable and informative different properties of the input are (e.g. MacWhinney, 1987; MacWhinney & Bates, 1989; MacDonald, 1993; Spivey-Knowlton & Tanenhaus, 1994; Taraban & McClelland, 1990; Trueswell &

Tanenhaus, 1994). In their extreme formulation, statistical accounts would allow for different solutions to different types of ambiguity even within the same language, as these accounts claim that attachment decisions are determined by corpus statistics.

All these models focus on the principles that govern attachment decisions in the face of ambiguous input, so they can be described as characterizing the control structure of the language architecture. They make claims about composed representations or attachments, for instance, by stating that a given attachment is preferred over the other or is computationally more costly to maintain than another. However, these accounts usually leave unspecified the other two components of the language architecture, the representation and processes. For instance, they assume the implementation of attachment decisions without specifying the underlying processes (e.g. memory retrieval, inhibition, spread of activation) that implement these attachments in real time. As such, most of the previous work on cross-linguistic variation can in principle be carried out without making commitments to the processes and memory structures that give rise to them.

In contrast, the focus of this thesis is to examine how cross-linguistic variation affects the nature of the representations and processes that speakers deploy during comprehension. This work can be seen as complementary to the models outlined above, as it aims to re-examine questions about the flexibility of the language architecture by making use of an explicit theory about the architecture of the memory system that is independently motivated by research on domain-general working memory (McElree, 2000; Gordon, Hendrick & Johnson, 2001; Gordon, Hendrick & Levine, 2002; McElree, Foraker & Dyer, 2003; Lewis & Vasishth, 2005; Lewis,

Vasishth & Van Dyke, 2006; Wagers, 2008). With this goal, I focus on a key process, memory retrieval, and I examine how systematic changes in grammatical features across languages lead to systematic changes in the process of memory retrieval and in the nature of the memory representations that are retrieved.

1.2 A cognitive architecture for retrieval

The process examined in this dissertation is memory retrieval, the ability to reaccess previous information in memory. This mechanism plays an important role in most current models of language processing, although its specific formulation can vary. Its generality is evidenced in the fact that retrieval operations have been proposed to subserve performance at the level of word and sentence processing and also in linguistic and non-linguistic tasks (for review, see Jonides et al., 2008). These facts suggest that retrieval is a pervasive feature of human cognition, and thus it is likely to be shared across languages.

I adopt the assumption that retrieval operates within an architecture in which only a limited amount of information can be simultaneously processed (McElree & Dosher, 1989; Lewis, Vasishth & Van Dyke, 2006; Wagers & Phillips, 2014). Since the nature of the linguistic input is incremental and sequential (i.e. words unfold one at a time), and only a small amount of information can be maintained in an active state, the processing of dependencies where the elements in a relationship can be separated by an unbounded number of words, namely *long distance dependencies*, necessitates retrieval. For example:

- (2) The **garden** in the outskirts of the city **was** littered with bottles and wrinkled papers.
- (3) The maintenance men told the **singer** about a problem. They had broken **his** piano and would have to repair that first.

In (2), when speakers reach the verb “was”, they need to check that it agrees in number features with the head of the subject phrase, “the garden”. As the activation of the head has decayed since its initial appearance and it is no longer available in the bottom-up input, speakers will need to retrieve it from memory. This involves using the cues in the current input (e.g. the number and person cues on the verb) to probe the stored pieces of information according to the content of their representations in a simultaneous and direct fashion, termed *direct access* (Anderson, 2005; Cowan, 2001; Lewis and Vasishth, 2005; Lewis, Vasishth, & Van Dyke 2006; McElree, 2006). In this process, other noun phrases previously encoded (“the outskirts”, “the city”) are accessed and evaluated as well. Sentence (3) illustrates a similar case with coreference: the pronoun “his” needs an antecedent matching in animacy, person, number and gender features. Retrieval in this case involves the parallel access of previously encoded nouns such as “maintenance men”, “singer” and “problem”, and the word that matches the most cues (“singer”) is the most likely to be retrieved.

Retrieval processes for long-distance dependencies such as pronouns and verbal agreement strongly rely on morphosyntactic features such as number, gender, person and animacy. That these features are not actively maintained across sentence boundaries was demonstrated very early on in psycholinguistics by studies on verbatim recall (Jarvella, 1971; Potter & Lombardi, 1998; Sachs, 1967). These

experiments showed that memory for “surface” properties such as syntactic information decays very quickly in processing. For example, Sachs (1967) presented participants with spoken texts containing sentences such as the one in (4):

- (4) He sent a letter about it to Galileo, the great Italian scientist.

After a variable time delay, participants were shown a target sentence and they had to judge whether it was identical to the sentence in the original text. Sachs varied the type of relationship between the original and the target sentence, as well as the amount of interpolated material, which was either 0, 80 syllables (approximately 27 sec) or 160 syllables (approximately 46 sec). The target sentence was either identical to the original sentence in the passage, or it contained (i) a semantic change such as reversed thematic roles (e.g. “Galileo, the great Italian scientist, sent him a letter about it”); (ii) a syntactically altered but truth-conditionally equivalent version of the original (e.g., “A letter about it was sent to Galileo, the great Italian scientist”); (iii) a change in surface form (e.g. “He sent Galileo, the great Italian scientist, a letter about it”). Sachs found that only when the target sentence differed in semantic content from the original sentence could participants reliably discriminate between them. After 80 syllables, participants were still good at remembering the meaning of the original sentence, but they were basically at chance with variations in the sentence’s syntactic structure, or in recognizing changes between active and passive form.

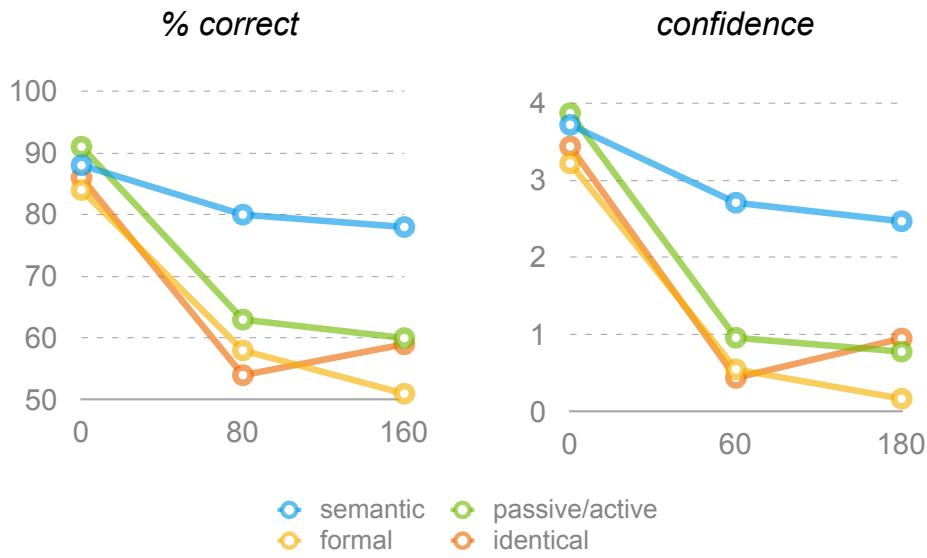


Figure 1 Results in Sachs (1967). *Left panel*: percent correct responses on the “identical vs. changed” judgment task. *Right panel*: participants’ self-reported confidence in their judgments. Reproduced from Wagers (2007).

These results suggested that structural and surface form information were quickly lost. In contrast, participants’ memory for semantic information was more durable. This second claim will be taken as another important property of the architecture of the language system in this thesis. I will assume a difference between linguistic representations, which contain morphosyntactic features and decay quickly in memory, and conceptual or discourse-based representations, which are more durable and encode conceptual information about the individuals, properties and events described by linguistic utterances. In a nutshell, discourse representations provide a description of participants’ understanding of the events described by a sentence.

One of the goals of this dissertation is to show that another source of cross-linguistic variation may be in the nature of the representations that speakers retrieve

across languages. I will show that in addition to previously found time-course differences in the decay of linguistic and discourse representations, retrieval processes can be shown to target representations of different kind across languages. I will propose that variation in the grammatical features of each language can determine whether linguistic or discourse representations (or both) are selected as targets for retrieval.

1.3 Two test cases of cross-linguistic variation

The previous section argued that parallel, cue-based memory retrieval is a key component of the processing of long-distance dependencies (see Chapter 2 for a similar argument for predictive processes). My goal is to examine how this retrieval mechanism interacts with language-specific features. However, there is no shortage of long-distance dependencies in language: reflexives, agreement, case, ellipsis, raising, thematic binding and wh-questions are among them. This thesis focuses on two dependencies, since their processing depends on features along which languages vary systematically: number agreement and coreference.

1.3.1 Agreement

Consider first the case of number representations. In English, verbal morphology only marks number in the third person singular (e.g. “You write” vs. “She writes”). In contrast, in languages like Spanish and Italian morphology is richer and more marked, as illustrated in (5). In addition, agreement is also functionally

more important in Spanish. One reason for this is that Spanish has a freer word order, so sentence initial position is not as reliable a cue to subjecthood as it is in English. Relatedly, Spanish is a null-subject language, which allows speakers to omit subjects in their utterances. These two facts make agreement information a necessary cue for speakers to infer subjecthood relationships.

- (5a) And yet the animals never gave up hope. More, **they** never lost, even for an instant, **their** sense of honor and privilege in being members of Animal Farm.
(George Orwell, *Animal Farm*)
- (5b) Y aún así, **los** animales nunca renunciaron a la esperanza. Y (**ellos**) nunca perdieron, ni siquiera por **un** instante, **sus** sentidos de honor y privilegio por ser miembros de **la** Granja de Animales.

In addition, the availability of number morphology has been shown to impact the acquisition of agreement across languages: children acquiring languages with richer morphological systems learn them faster and make less agreement errors in development (Slobin, 1985; Phillips, 2010; Pinker, 1984). This is shown in the graph below, which illustrates the rate of production of root infinitives (i.e. cases of missing agreement morphology on the verb such as “Patsy need a screw”) as a function of age. Children exposed to morphologically richer languages use fewer root infinitives and emerge from the root infinitive stage at a younger age.

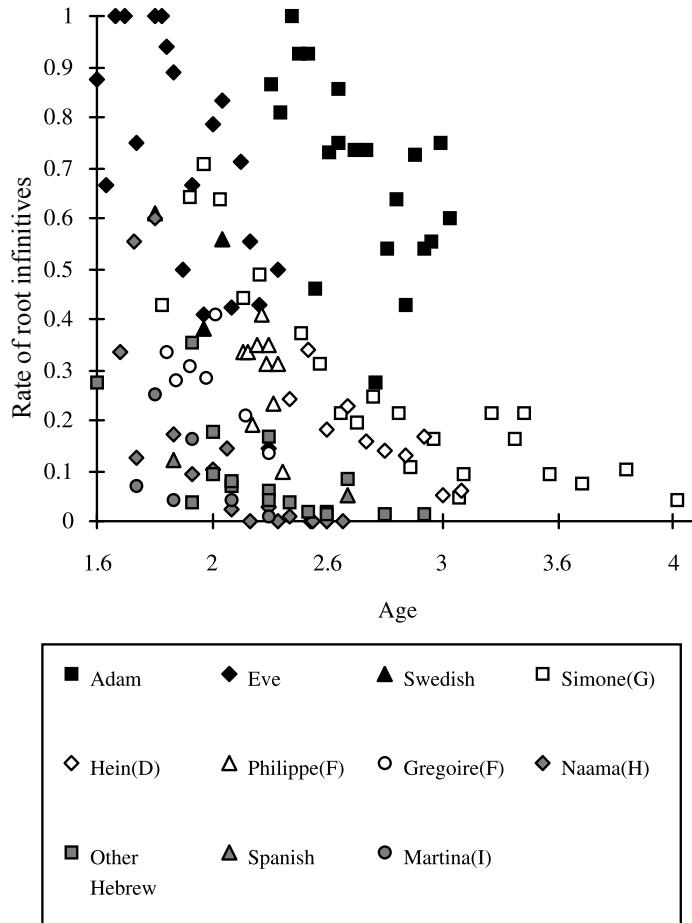


Figure 2 Use of root infinitives across languages. Children are separated into 3 groups according to the richness of the inflectional paradigm in their language. Black markers represent English and Swedish, which distinguish no more than two forms in any tense. Proportions of root infinitives are highest among children in this group. Unfilled markers represent French, German, and Dutch, which distinguish up to 3 or 4 forms for regular main verbs in any tense. Proportions of root infinitives are markedly lower for this group. Finally, grey markers represent Spanish, Catalan, Hebrew, and Italian, which show the richest agreement paradigms. Proportions of root infinitives are lowest in this group. Reproduced from Phillips (2010).

These facts show that variation in the richness of number information is relevant formally and in processing. Formally, it can be related to other syntactic properties of a language such as its pro-drop status and word order constraints. In processing, number morphology varies in functional relevance for adult speakers of different languages and it correlates with differences in the speed of agreement

acquisition for child learners. In the domain of agreement, the question asked in this dissertation is: *do differences in the availability of number morphology across languages drive differences in the strength of the number representations that speakers retrieve from memory?* In Chapter 2 I examine whether the richness of number morphology in Spanish results in stronger number representations in memory, and whether this causes differences in how retrieval of number information takes place across languages. I address this question by comparing the processing of agreement errors across Spanish and English comprehension.

1.3.2 Coreference

The second test case analyzed here is the processing of pronouns during coreference. In this case, the feature that differs across languages is the availability of syntactic gender. In English, gender is either semantic (e.g. “boy”, “king”) or stereotypical (“janitor”, “miner”). In addition, most morphosyntactic and lexical features of English nouns that are relevant for coreference such as animacy, number and gender have conceptual correlates, and thus it is reasonable to assume that they can also be represented at the discourse level. This has led to the proposal that referential pronouns in English only access the referent of their antecedent in a discourse model (the *discourse-only hypothesis*; Cloitre & Bever, 1988; Lucas, Tanenhaus & Carlson, 1990; Sag & Hankamer, 1984).

In languages with syntactic gender, on the other hand, gender is an arbitrary property of nouns and often lacks conceptual correlates. This is most evident with inanimate nouns: for example, in Spanish the gender difference between the nouns

“meat” (which is feminine) and “chicken” (masculine), which appear in sentence (6), is marked morphologically but has no conceptual consequences. Therefore, there is no motivation for including syntactic gender information at the discourse level:

- (6) Anabel volvió del supermercado y dejó **el pollo** sobre la mesa y **la carne** en la heladera. A la noche, su novio **lo** usó para cocinar la cena.

Annabel came back from the supermarket and left the chicken_{masc} on the table and the meat_{fem} in the fridge. In the evening, her boyfriend used it_{masc} to make dinner.

Any speaker of Spanish interprets the pronoun “it” in (6) as referring to the noun phrase “the chicken”, and not “the meat”. This interpretation can only be obtained from the agreement in gender between “it” and “chicken”, since both nouns denote otherwise equally plausible elements for dinner. In processing, previous studies have found that speakers of languages with syntactic use gender agreement between pronouns and antecedents as a cue for disambiguation in comprehension (Cacciari, Carreiras & Barbolini-Cionini, 1997; Carreiras, Garnham & Oakhill, 1993; Frazier, Henstra & Flores d' Arcais, 1996; Garnham, Oakhill, Erlich & Carreiras, 1995) and pronoun selection in production (Meyer & Bock, 1999). These findings have suggested that in languages with syntactic gender, morphosyntactic information of a pronoun’s antecedent in the lexicon must be reactivated during coreference in addition to its referent in the discourse.

The previous facts suggest that while English speakers might retrieve only the referent in their discourse model, speakers of languages with syntactic gender need to retrieve both lexical and discourse level representations. Chapters 3 and 4 ask: *what types of antecedent representations do speakers of languages with syntactic gender*

retrieve? Do these representations differ from those retrieved by English speakers? I put forth a specific account of how the existence of syntactic gender can drive differences in the type of representations that speakers retrieve during coreference. This account is examined in two eye-tracking studies comparing German, a language with syntactic gender, with English, a language that lacks it.

1.4 Outline of the dissertation

Chapter 2 examines whether there are cross-linguistic differences in the processing of number agreement. It addresses whether differences in the availability of number morphology drives differences in the strength of number representations across languages. Previous studies have found that English speakers experience attraction effects when comprehending subject-verb agreement, showing eased processing of ungrammatical sentences that contain a syntactically unlicensed but number matching attractor noun. In three self-paced reading experiments I examine whether attraction effects also occur in Spanish, a language where agreement morphology is richer and functionally more significant. I find that despite having a richer morphology, Spanish speakers show reliable attraction effects in comprehension, and that the magnitude and distributional profile of these effects are strikingly similar to those found in English. Further, I show that attraction profiles in Spanish but not English vary as a function of the type of verb that carries agreement information. Spanish speakers behave like English speakers when agreement information is associated with main verbs, but a different behavior is observed with auxiliary verbs. I consider different views to account for the pattern observed with

Spanish auxiliary verbs and conclude that a predictive account offers the best coverage. I propose that in comprehension, differences in morphological richness between languages may impact prediction to a greater extent than retrieval processes.

Chapter 3 motivates the existence of cross-linguistic differences in the processing of coreference. I outline the arguments used to claim that coreference with pronouns in English is established at the level of discourse. I examine the properties of pronominal coreference in languages with syntactic gender and show that, in contrast with English, these properties require access to more than a discourse-based representation of an antecedent noun. I then review experimental findings that support the processing claim that a *lexical* representation of a pronoun's antecedent must be reaccessed. I argue that if previous findings from languages with syntactic gender are taken together with the English findings, these data suggest that different antecedent representations must be retrieved across languages. In the following sections I review a previous experimental result that provided the main evidence against lexical re-access in English. I evaluate several potential concerns about the design of this experiment, and present two eye-tracking experiments in English that were conducted to address these issues.

Chapter 4 also focuses on the nature of antecedent representations reactivated during coreference. I present a new paradigm to study antecedent retrieval during reading comprehension, and present two eye-tracking studies in English and German, which examine whether there is differential reactivation of syntactic and semantic information. I find that German, but not English speakers, show semantic effects during pronoun processing, and argue that this lends support to the claim that

pronouns' antecedents may be reactivated at different levels of representation across languages. Specifically, antecedents in English may only be retrieved at the level of discourse, whereas German speakers may retrieve a lexical representation in addition to the conceptual one.

In the final chapter, I synthesize the empirical findings reported in this dissertation and discuss their implications for the study of cross-linguistic variation more broadly. I discuss how to enrich current theoretical descriptions of predictive and retrieval computations, and outline ideas for future research.

2 Agreement processing in a morphologically rich language

2.1 Outline

Chapter 2 examines the processing of agreement in Spanish, a language with rich verbal morphology. Previous studies have found that English speakers experience attraction effects when comprehending subject-verb agreement, showing eased processing of ungrammatical sentences that contain a syntactically unlicensed but number matching attractor noun. In three self-paced reading experiments we examine whether attraction effects also occur in Spanish, a language where agreement morphology is richer and functionally more significant. We find that despite having a richer morphology, Spanish speakers show reliable attraction effects in comprehension, and that the magnitude and distributional profile of these effects are strikingly similar to those found in English. Further, we show that attraction profiles in Spanish but not English vary as a function of the type of verb that carries agreement information. Spanish speakers behave like English speakers when agreement information is associated with main verbs, but a different behavior is observed with auxiliary verbs. We consider different views to account for the pattern observed with Spanish auxiliary verbs and we conclude that a predictive account offers the best coverage. We propose that in comprehension, differences in morphological richness between languages may impact prediction to a greater extent than retrieval processes.

2.2 Introduction

Languages differ widely in the degree to which their morphological systems are used to convey formal and conceptual distinctions. One important unanswered question is: what impact do these properties have on core language processing mechanisms? For example, the degree to which speakers rely on morphological cues in comprehension may critically depend on the extent to which these cues are made available by a language's morphology (MacWhinney, 1987; MacWhinney, Bates, & Kliegl, 1984). This could have important consequences for the processing of grammatical relationships such as agreement: agreement errors in comprehension might be more common in languages with an impoverished morphology, and particular challenges might arise for learners of a morphologically richer second language (Jia, Aaronson, & Wu, 2002; Jiang, 2004, 2007; McDonald, 2000).

In the current study we investigate the question of how morphological variation impacts comprehension mechanisms by focusing on two key cognitive capacities central to language comprehension: the ability to retrieve previous information from memory (Caplan & Waters, 2013; Gordon, Hendrick, & Johnson, 2001; Lewis & Vasishth, 2005; McElree, Foraker, & Dyer, 2003) and the ability to make predictions about what material will be encountered next (Altmann, van Nice, Garnham, & Henstra, 1998; Kazanina, Lau, Lieberman, Yoshida, & Phillips, 2007; Lau, Stroud, Plesch, & Phillips, 2006; Staub & Clifton, 2006; Yoshida, Dickey, & Sturt, 2012). We examine the comprehension of subject-verb agreement, which is likely to involve both prediction and retrieval; number features on the verb can be predicted from the subject noun, while number features on the subject noun may also

be retrieved from memory when a verb is encountered. If cross-linguistic variation impacts the operations used in comprehension, it could in principle affect prediction, retrieval or both. We explore these possibilities and we ask: are prediction and retrieval implemented uniformly across languages, or does their use vary according to the properties of each language?

We address this question by comparing the computation of subject-verb agreement in Spanish and English. In English, number morphology is limited, so word order and syntactic information are the most reliable cues to resolve subject-verb dependencies (MacWhinney, Bates, & Kliegl, 1984; Severens, Jansma, & Hartsuiker, 2008). In Spanish, however, number agreement obeys the same syntactic requirements as in English but morphology is both more marked and functionally more significant. This is illustrated in the two passages below (agreement morphology is bolded):

- (1a) And yet the animals never gave up hope. More, **they** never lost, even for an instant, **their** sense of honor and privilege in being members of Animal Farm. (George Orwell, *Animal Farm*)
- (1b) Y aún así, **los** animales nunca renunciaron a la esperanza. Y (**ellos**) nunca perdieron, ni siquiera por **un** instante, **sus** sentidos de honor y privilegio por ser miembros de **la** Granja de Animales.

These passages highlight several differences between Spanish and English. First, Spanish morphology marks number agreement more strongly. This is the case in nominal phrases, where the nouns and all of their modifiers carry agreement information in Spanish (e.g. “sus sentidos”, “los animales”), while in English plural number is mainly marked on the head noun. This is also the case in verbal phrases, where Spanish verbs mark agreement for all syntactic persons, and where plural verb

forms often differ sharply from the singular forms (e.g. “renunció vs. “renunciaron”). Second, agreement morphology is functionally more important in Spanish. One reason for this is that Spanish has a freer word order, so sentence initial position is not as reliable a cue to subjecthood as it is in English. Relatedly, Spanish is a null-subject language, which allows speakers to omit subjects in their utterances (e.g. “(ellos) perdieron”). These last two properties make verbal morphology necessary to identify the subject of a sentence, and this has given rise to the claim that morphological cues are more reliable than positional information in Spanish (Kail, 1989; MacWhinney, 2001).

In the current study we investigate whether the increased richness and reliability of morphological cues in Spanish relative to English impact retrieval and prediction mechanisms in the computation of subject-verb agreement. Cross-linguistic variation is expected under frameworks like the Competition Model (e.g. MacWhinney, 1987; MacWhinney & Bates, 1989), where more reliable cues are predicted to be acquired first and to most strongly affect language comprehension processes. Therefore, if Spanish comprehenders rely more on morphological cues than English comprehenders, they might be less vulnerable to agreement errors. In the rest of the Introduction, we first summarize previous findings of agreement attraction errors in English. We describe a retrieval mechanism that has been proposed to give rise to these errors. We then discuss how this mechanism might be used differently in Spanish, and present an overview of three experiments that were carried out to address the relationship between morphological richness and agreement computations.

Agreement attraction in comprehension

The processing of agreement is known to be prone to errors. A particular type of error called *agreement attraction* occurs when the verb in a sentence agrees not with its grammatical subject but with a number-marked noun in its vicinity, conventionally called an “attractor”. Errors are especially common when the grammatical subject is singular and the attractor is plural. In comprehension, plural attractors have been shown to reduce processing difficulty in ungrammatical sentences (Dillon, Mishler, Sloggett, & Phillips, 2013; Nicol, Forster, & Veres, 1997; Pearlmuter, Garnsey, & Bock, 1999; Wagers, Lau, & Phillips, 2009). For example, given the ungrammatical sentence “The key to the cabinet(s) are on the table”, comprehenders read the words following the ungrammatical plural verb more quickly when the attractor matches the verb in number (“cabinets”) than when it does not (“cabinet”). This suggests that ungrammatical sentences are associated with reduced processing cost when they contain a syntactically unlicensed but number-matching attractor.

Here we investigate the view that processing errors arise due to properties of memory retrieval (Anderson, 1983; Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke 2006; McElree, 2006; Ratcliff, 1978). According to this mechanism, when a word in a syntactic dependency is encountered, it triggers a retrieval to find its dependent element. Retrieval is accomplished by a simple memory access mechanism that uses content-based cues to directly retrieve items in memory, and the retrieval cues are provided by the word. For example, in the case of subject-verb agreement,

the retrieval cues provided by the verb are syntactic, semantic and morphological. The preceding words in the sentence are queried in parallel according to these cues, and the word that matches the most cues is the most likely to be retrieved.

Wagers, Lau & Phillips (2009, henceforth WL&P) provided one key finding in favor of a cue-based retrieval explanation for agreement attraction. They used relative clause constructions (RCs), where the plural attractors did not intervene linearly between the critical subject-verb pair (see also Clifton, Frazier, & Deevy, 1999; Staub, 2009, 2010). They manipulated sentence grammaticality and attractor number using a self-paced reading paradigm (Experiment 2). Grammaticality was varied by making the RC verb agree (2a-b) or disagree (2c-d) with the RC subject, “reviewer”. The attractor noun was outside the RC, and it was in either the singular or plural form (“musician(s)”):

- (2a) The musician [who the reviewer praises so highly] will probably win a Grammy.
- (2b) The **musicians** [who the reviewer praises so highly] will probably win a Grammy.
- (2c) *The musician [who the reviewer praise so highly] will probably win a Grammy.
- (2d) *The **musicians** [who the reviewer praise so highly will probably win a Grammy.

WL&P found that a plural attractor affected reaction times only in the ungrammatical conditions, where the post-verbal region in the RC was read more quickly when the attractor was plural (2d) than when it was singular (2c). In contrast,

in the grammatical conditions the attractor did not affect reaction times. The authors described this pattern as the “grammatical asymmetry”, and suggested that this asymmetry is expected if retrieval is used as a repair or reanalysis mechanism as the consequence of a violated number prediction. On this view, the number of the verb is predicted at the subject noun within the RC. When the verb form violates this prediction (e.g. “praise”), participants use a cue-based retrieval to check whether the correct feature was somehow missed during the first pass. The cues used in the search are provided by the verb, and consist of morphological cues like [Number: Plural], syntactic cues like [Role: Subject], and clause-bounding cues (see Wagers, 2008 for discussion). In the ungrammatical condition with a plural attractor, the outcome of this search is a scenario in which the two previous nouns partially match the retrieval cues. The RC subject “reviewer” agrees with the verb in syntactic (but not morphological) cues, while the attractor “musicians” agrees in morphological (but not syntactic) cues. Due to this partial match, the attractor is sometimes mistakenly retrieved. When this occurs, agreement can be established, facilitating processing in (2d) as compared with (2c), where agreement always fails. In contrast, in the grammatical conditions the verb always matches the number prediction made by the subject noun, and therefore retrieval is not engaged and no attraction takes place.

In contrast, other accounts of agreement attraction would predict symmetric attraction effects, that is, attraction effects in both grammatical and ungrammatical sentences. For example, percolation accounts posit that attraction results from faulty movement or “percolation” of plural features from the attractor to the subject noun (Eberhard, Cutting, & Bock, 2005; Franck, Vigliocco, & Nicol, 2002; Pearlmutter,

Garnsey, & Bock, 1999). This causes the plural feature of the attractor to sometimes overwrite the number of the subject phrase. When this happens, speakers should construct an incorrect representation of the number of the subject phrase and symmetric attraction effects should occur: grammatical sentences with a plural attractor and a singular verb (2b) should sometimes be perceived as ungrammatical (yielding processing difficulty), while ungrammatical sentences with a plural attractor and a plural verb (2d) should sometimes be perceived as grammatical (yielding facilitation). This prediction of the percolation account appears inconsistent with the asymmetric attraction effect that WL&P obtained.

Although asymmetric attraction effects cannot be explained by a percolation account, it is worth noting that symmetric attraction might be expected under a strong version of the predictive mechanism in WL&P's account. These authors only considered the effect of the number predictions initiated by the subject noun within the RC. However, the attractor noun, which is the main clause subject, should also lead comprehenders to make a prediction for the number of the main clause verb. If these predictions can interfere with each other, they might create processing difficulty at the singular RC verb in the grammatical attractor sentences (2b) since the number of the verb matches the RC subject prediction but not the main clause subject prediction. Although WL&P's studies found no evidence of attraction in grammatical sentences, a language with a richer morphology like Spanish might show differences not only in retrieval but also in predictive processes such as the generation of number expectations. Therefore, in this study we re-examine the existence of asymmetric effects, and we contrast the predictive mechanism in WL&P's account with

percolation accounts. Distinguishing between these accounts is important because although they both allow for symmetric attraction effects, they make very different assumptions about the underlying architecture of the language processing system. While percolation accounts assume that the processing system can construct incorrect or ungrammatical representations, WL&P's account does not, and it attributes symmetric attraction effects to the conflict between the number of the RC verb and two different number predictions generated by the previous nouns in the sentence.

The present study

The goal of the present study was to determine whether Spanish, a morphologically rich language, shows agreement attraction effects in comprehension. To our knowledge, while there have been several studies on agreement violations in morphologically rich languages (Barber & Carreiras, 2005; see Mancini, Molinaro, & Carreiras, 2013, for review), no published comprehension studies have looked at subject-verb agreement attraction in a language other than English. In contrast, there have been numerous cross-linguistic studies on attraction effects in production (Bock, Carreiras, & Meseguer, 2012; Franck, Vigliocco, Antón-Méndez, Collina, & Frauenfelder, 2008; Lorimor, Bock, Zalkind, Sheyman, & Beard, 2008; Vigliocco, Butterworth, & Garrett, 1996; Vigliocco, Butterworth, & Semenza, 1995; Vigliocco, Hartsuiker, Jarema, & Kolk, 1996; Hartsuiker, Schriefers, Bock, & Kikstra, 2003). However, the role of morphological richness in these studies remains unresolved, as some have found higher attraction rates in languages with poor morphology, whereas

others have found the opposite pattern (see Lorimor, Bock, Zalkind, & Sheyman, 2008, for discussion).

We consider several ways in which richer morphology could affect the retrieval mechanism, causing Spanish and English speakers to behave differently. One possibility, as mentioned above, is that Spanish speakers could be less susceptible to attraction than English speakers. If retrieval is only engaged as a reanalysis strategy, it might be attempted less frequently in a language with richer morphology. In English, number morphology is used sparingly in nominal phrases: for example, in WL&P's experiment, only the attractor noun carried a plural marking. Therefore, English comprehenders might have been less certain about their encoding of the previous input, making them more likely to attempt reanalysis and resulting in increased susceptibility to attraction. In Spanish, in contrast, plural number is clearly marked on both the determiner and attractor noun, and plural verb forms are different than the bare forms. Therefore, Spanish comprehenders might be more certain about their encoding of the previous input such that they attempt reanalysis less often, resulting in lower or non-existent attraction rates in comprehension in Spanish.

In contrast, a second possibility that is more closely related to theoretical accounts of cue usage is that richer verb morphology could in fact make Spanish comprehenders more susceptible to attraction than English comprehenders. If richer morphology affects retrieval through differential cue weighting, and if morphological cues are weighted more heavily in Spanish, then comprehenders might make more partial-match errors at retrieval, resulting in larger attraction rates. This possibility is motivated by the fact that verbal agreement morphology is more necessary in Spanish

due to its freer word order and null subject status, and therefore, morphological cues might carry more weight at retrieval relative to syntactic cues, making spurious retrieval of the attractor noun more likely.

We contrast these two possibilities with the hypothesis that the retrieval mechanism is used uniformly across languages. If retrieval is implemented similarly cross-linguistically, and if subject and verb number are mentally represented similarly across languages, regardless of surface form, then similar attraction rates should be found in English and Spanish. Under this hypothesis, we expect Spanish speakers to show reduced processing cost for ungrammatical sentences when they contain a plural attractor. Further, we expect to also observe a grammatical asymmetry in Spanish, yielding an attraction effect in ungrammatical sentences but no attraction effect in grammatical sentences. Overall, observing qualitatively and quantitatively similar attraction effects in Spanish and English would suggest that retrieval is implemented uniformly across the two languages.

We also tried to obtain a more detailed analysis of the time course of attraction effects using distributional analyses. These analyses have been productively applied by two previous studies on agreement attraction using a forced-choice paradigm (Staub, 2009, 2010). These studies found that attraction effects in relative clauses result from a process that involves only a small proportion of the trials, and it is thus seen in the right tail of a reaction time distribution. We wanted to examine whether a similar distributional pattern was obtained when using a reading paradigm that did not require participants to make conscious choices between alternative verb forms. Secondly, our goal was to address a time course prediction associated with

WL&P's account, which posits that comprehenders engage retrieval as a repair mechanism after they detect a number mismatch between the verb and the subject within the RC. Under this error-driven account of retrieval, grammaticality violations should impact reaction time distributions earlier than attraction effects, as attraction effects reflect the output of a retrieval process that follows the initial detection of a violation. In order to test this prediction, we used vincentile plots to compare the temporal onsets of agreement violations and attraction effects.

Overview of the experiments

Three comprehension experiments were conducted using a self-paced reading paradigm. Experiments 1 and 2 addressed whether Spanish comprehenders are susceptible to attraction effects, and whether their online reading profile is similar to English. Since we were interested in assessing the effect of morphological richness in agreement attraction, we also varied the markedness of the verb forms. Previous English experiments with relative clauses used main verbs that had singular and plural forms differing in only one character (e.g. “praises vs. praise”). In Experiment 1, we used main verbs in Spanish, which contrasted more sharply in the singular and plural forms. In Experiments 2 and 3 we used auxiliary verbs in Spanish and English, which are more closely matched across languages in length between singular and plural-marked verbs. In all three studies we complement the mean reading time results with an analysis of the reaction time distributions of attraction effects.

2.3 Experiment 1: Spanish Main Verbs

The goal of Experiment 1 was to investigate the properties of number attraction in Spanish comprehension. If the same retrieval mechanism underlies attraction across languages, then we expect English and Spanish comprehenders to display similar reading-time profiles, with plural attractors facilitating processing only in the ungrammatical conditions, as shown in English by WL&P. Furthermore, if retrieval is deployed as a repair mechanism when comprehenders detect number violations, then we expect a specific time course profile: grammatical violations should affect reaction time distributions earlier than attraction effects. In order to address this prediction, and to also replicate the previous observation that attraction effects in relative clauses result in a change of the right tail of reaction time distributions (Staub 2009, 2010), we performed distributional analyses using a non-parametric technique, vincentile plots, explained further below.

2.3.1 Methods

Participants

Participants ($n = 32$, mean age = 27 years, 15 females) were all native speakers of Argentinian Spanish and were recruited from the University of Buenos Aires community. All participants provided informed consent and were compensated the equivalent of \$5/hour.

Materials and design

The design was identical to the one used in Experiment 2 of WL&P and the Spanish materials were constructed based on their items. They consisted of 48 sentence sets arranged in a 2×2 within-subjects design, with grammaticality (*grammatical/ungrammatical*) and attractor number (*singular/plural*) as factors. For the grammaticality factor, we manipulated the agreement relationship between the subject noun and the verb within the relative clause: in the grammatical conditions, the subject and the relative clause verb were both singular (i.e., they agreed in number), while in the ungrammatical conditions the subject was singular and the verb was plural (i.e., they mismatched in number) which rendered the sentences ungrammatical. The relative clause verbs were in the simple past tense and perfective aspect. The singular suffix for this tense-aspect combination in Spanish is one character long (e.g. “escribi-ó”, *read.3sg*), while the plural suffix is 4 characters long (e.g. “escribi-eron”, *read.3pl*). To allow for the presence of spillover effects (Just & Carpenter, 1978) an adverbial prepositional phrase consisting of two or three words was introduced after the relative clause verb.

The head noun modified by the relative clause was considered the “attractor”. The attractor noun was always inanimate (e.g. “nota”, *note*) and could be either singular or plural. In contrast, the subject noun within the relative clause was always singular and animate (e.g. “chica”, *girl*). The 48 sentence sets were distributed across four lists in a Latin Square design, and were combined with 24 items (half ungrammatical) from a different experiment not reported here, and 188 grammatical filler sentences of similar length. This resulted in 13.8% of the items being

ungrammatical. An example set of experimental materials is presented in Table 1 and the full item set is available in Appendix A.

Spanish main verb items (Experiment 1)	
<i>Gram, sg attractor</i>	La nota que la chica <u>escribió</u> en la clase alegró a su amiga. <i>The note that the girl wrote_{sg} during class cheered her friend up.</i>
<i>Gram, pl attractor</i>	Las notas que la chica <u>escribió</u> en la clase alegraron a su amiga. <i>The notes that the girl wrote_{sg} during class cheered her friend up.</i>
<i>Ungram, sg attractor</i>	* La nota que la chica <u>escribieron</u> en la clase alegró a su amiga. <i>The note that the girl wrote_{pl} during class cheered her friend up.</i>
<i>Ungram, pl attractor</i>	* Las notas que la chica <u>escribieron</u> en la clase alegraron a su amiga. <i>The notes that the girl wrote_{pl} during class cheered her friend up.</i>
Spanish auxiliary verb items (Experiment 2)	
<i>Gram, sg attractor</i>	La nota que la chica <u>ya</u> a escribir en la clase alegrará a su amiga. <i>The note that the girl is going to write during class will cheer her friend up.</i>
<i>Gram, pl attractor</i>	Las notas que la chica <u>ya</u> a escribir en la clase alegrarán a su amiga. <i>The notes that the girl is going to write during class will cheer her friend up.</i>
<i>Ungram, sg attractor</i>	* La nota que la chica <u>van</u> a escribir en la clase alegrará a su amiga. <i>The note that the girl are going to write during class will cheer her friend up.</i>
<i>Ungram, pl attractor</i>	* Las notas que la chica <u>van</u> a escribir en la clase alegrarán a su amiga. <i>The notes that the girl are going to write during class will cheer her friend up.</i>

Table 1 Sample set of experimental items in Experiments 1 and 2.

Procedure

Sentences were presented on a laptop PC using the Linger software (Doug Rohde, MIT) in a self-paced word-by-word moving window paradigm (Just, Carpenter, & Woolley, 1982). Each trial began with a screen presenting a sentence in which the words were masked by dashes. Each time the participant pressed the space bar, a word was revealed and the previous word was re-masked. The time spent on each word was measured as the time difference between two successive key presses. Every experimental and filler item was followed by a yes/no comprehension question to ensure that participants were attending to the stimuli. The comprehension questions

never referred to the agreement dependency. On-screen feedback was provided for incorrect answers. Participants were instructed to read at a natural pace and answer the questions as quickly and accurately as possible. Three practice items were presented before the beginning of the experiment. In all of the self-paced reading experiments reported here, participants were never informed that sentences would contain grammatical errors. The order of experimental and filler items was randomized for each participant. The entire experimental session lasted approximately one hour.

Analysis

Comprehension questions

A statistical analysis of the proportion of correct responses in the experimental trials was carried out using mixed effects logistic regression (Jaeger, 2008), with grammaticality, attractor number and their interaction as fixed effects. For consistency, the random effects structure of this model was identical to the one used in the reaction time analysis (see below). Analyses were carried out using R, an open source programming language and environment for statistical computing (R Development Core Team, 2014), and in particular the *lme4* package for linear mixed effects models (Bates, Maechler, & Bolker, 2013). Only data from participants with accuracy above 70% in the filler items were included in the analysis.

Reaction times

To allow for comparison with the English study by WL&P our analysis was maximally similar to theirs. Reaction times (RTs) that exceeded a threshold of 2.5 standard deviations by region and condition were excluded (Ratcliff, 1993). Across all self-paced reading experiments reported in this paper, on average this resulted in the exclusion of 2.4% of the trials in the critical regions (min: 2.2%, max: 3.1%). Also based in the previous studies, we identified two regions of interest: the verb in the relative clause (*verb region*) and the word immediately following it (*verb+1 region*).

The model used to analyze RTs included grammaticality, attractor number and their interaction as fixed effects. Both main effects were coded using orthogonal contrasts: for the grammaticality factor, the mean of the ungrammatical conditions was compared with the mean of the grammatical conditions; for the attractor number factor, the mean of the plural noun conditions was compared with the mean of the singular noun conditions. Following current guidelines in the psycholinguistics literature (Barr, Levy, Scheepers, & Tily, 2013), we initially constructed a maximal model that included random intercepts and slopes for all fixed effects and their interaction. However, as this maximal model failed to converge in most cases, we gradually simplified the random effects structure following the suggestions by Barr and colleagues. The results that we report correspond to the model with the maximal random effects structure that converged for all critical regions across experiments, which included by-subject random intercepts and slopes for both fixed effects, a by-item intercept and no correlation parameters between them. The model estimates in milliseconds ($\hat{\beta}$) are presented in the tables below, with negative values

corresponding to a decrease in RTs. Due to difficulty in generating a Markov chain Monte Carlo sample from the posterior distribution of the parameters, the current version of the *lme4* package no longer computes p-values (Bates, Maechler, & Bolker, 2013). Therefore, we provide t-values and 95% confidence intervals in milliseconds (CI) for all model estimates. Estimates with $|t| > 2$ and $|z| > 2$ are considered significant effects at the .05 level (Baayen, Davidson, & Bates, 2008; Vasishth, Brüssow, Lewis, & Drenhaus, 2008). In order to ensure consistency with WL&P, we also performed a complementary analysis computing ANOVAs by subjects (F_1) and items (F_2). This analysis yielded similar results to the linear mixed effects model and is presented in Appendix B.

Finally, one difference between WL&P's experiment and the current study is that due to Spanish verbal morphology, the ungrammatical verbs were on average three characters longer than the grammatical verbs. This is problematic, as both ungrammaticality and word length have been shown to increase reaction times. Therefore, in order to avoid a confound between these two factors, the effect of word length was estimated from the entire dataset, and it was then regressed from the raw reaction times using a linear model (cf. Hofmeister, 2010). Only the length-regressed reaction times were entered into the statistical analysis.

Vincentile plots

We constructed vincentile plots to examine the effect that each experimental factor had on the reaction time distributions (Ratcliff, 1979; Vincent, 1912). Vincentile plots are a non-parametric way of assessing the shape of a reaction time

distribution. They are plots of quantiles, estimated in a way that is robust to outliers and supports averaging across participants. They were constructed as follows. First, as the entire RT distribution was of interest, a conservative trimming procedure was used, excluding only RTs that exceeded 4000 ms (cf. Staub, 2010). This resulted in the exclusion of 0.1% of the trials in the critical regions (min: 0%, max: 0.4%) across the three experiments reported in this paper. Furthermore, only participants who had at least 10 trials per condition were included in the analysis. The raw RTs for each participant in each condition were binned into the shortest 10% (vincentile 1), the next shortest 10% (vincentile 2), etc. The mean of the observations in each vincentile was calculated. Finally, vincentile plots were computed by collapsing the same bins across subjects. These values were displayed as connected points on a plot with vincentile number on the *x*-axis and reaction time in milliseconds on the *y*-axis.¹

In order to compare grammaticality and attraction effects we computed three difference scores: a *grammaticality contrast* (computed as the subtraction of the grammatical from the ungrammatical vincentile curve in the no attractor conditions), an *attractor ungrammatical contrast* (obtained by subtracting the plural from the singular curve in the ungrammatical conditions) and an *attractor grammatical contrast* (obtained by subtracting the plural from the singular curve in the grammatical conditions). Therefore, for each participant, these difference scores consisted of a set of ten values for each contrast and they expressed the effect of

¹ One possible concern is that due to the number of trials per participant per condition in the experiment ($n = 12$) the use of 10 vincentiles involved having on average one data point per participant per bin, which is less than the number of observations traditionally used in vincentile analyses. To address this concern, we conducted a complementary analysis using only 4 vincentiles, which allowed having 3 observations per participant per bin. As the two analyses yielded very similar results, we present the analysis based on 10 vincentiles and we note any discrepancies in the text.

grammaticality, attraction in the ungrammatical conditions and attraction in the grammatical conditions respectively.

To examine the time course of each experimental factor, we then constructed a linear model with vincentile number as a categorical predictor. For each contrast, we compared the difference score in each vincentile with the mean difference score in the previous vincentiles (e.g., the difference score in the second vincentile was compared with the difference score in the first vincentile, the difference score in the third vincentile was compared with the mean difference score of the first and second vincentiles, etc.). This allowed us to define the divergence point as the earliest vincentile with a difference score significantly different from the previous vincentiles. All statistical analyses were performed using a model with the maximal random effects structure that converged for all critical regions across experiments. The final model included fixed effects of vincentile number, grammaticality (for the *grammaticality* contrast), attractor number (in the *attractor ungrammatical* and *attractor grammatical* contrasts), and by-subject random intercepts. As all vincentiles following the divergence point also displayed significant differences, we only report the statistics associated with the earliest vincentile where the difference became significant.

It is important to clarify that in previous sentence comprehension studies (Staub, 2011; Staub, White, Drieghe, Hollway, & Rayner, 2010) vincentiles have been often used to complement parametric analyses such as the fitting of an ex-Gaussian distribution to reading time data. In this study, however, due to the restricted number of trials per participant per condition ($n = 12$), we did not conduct

parametric analysis of the reaction time profiles. In contrast with parametric techniques, the use of vincentiles does not involve making assumptions about the underlying distribution of the observed reaction time data. Therefore, the vincentile plots are provided here with the goal of (i) illustrating the effect of each experimental factor on the shape of the reaction time distributions, and (ii) identifying the earliest point where each experimental effect impacted participants' reaction times.

2.3.2 Results

In the comprehension questions, mean accuracy was 95.1% and all participants had accuracy rates above 70%. The results from the logistic regression yielded a main effect of attractor number, with accuracy in the plural attractor conditions being slightly higher than in the singular conditions ($\hat{\beta} = 0.79$, CI = [0.08,1.50], $z = 2.17$).

The region-by-region reaction time averages are presented in Figure 1. Although the statistical analysis was always performed on the length-regressed RTs, we plot the raw RTs for easier readability. We analyzed the residual RTs in the two regions of interest, the verb region and the verb+1 region. At the verb region there were no main effects or interactions. The lack of an effect of grammaticality contrasts with the visual inspection of the plot, where the ungrammatical conditions show longer RTs than the grammatical conditions. However, after length was regressed from the RTs the effect of grammaticality was no longer significant at the verb region, suggesting that the difference seen in the plot was mostly due to the increased length of the verb forms in the ungrammatical conditions.

At the verb+1 region, there was a main effect of grammaticality, a main effect of attractor number and, crucially, a significant interaction between them. The main effect of grammaticality was due to the ungrammatical conditions having longer RTs than the grammatical conditions. The main effect of attractor number was due to the plural attractor conditions having shorter RTs than the singular noun conditions. Finally, the interaction shows that the number of the attractor noun affected grammatical and ungrammatical sentences differently. Pairwise comparisons indicated that attractor number had an effect in the ungrammatical sentences only, where the plural attractor condition was read significantly faster than the condition with a singular noun ($\hat{\beta} = -39$, CI = [-69,-11], $t = -2.75$). In the grammatical conditions, the number of the attractor had no significant effect ($\hat{\beta} = -6$, CI = [-20,8], $t = -0.80$). Table 2 shows the results from the mixed-effects model.

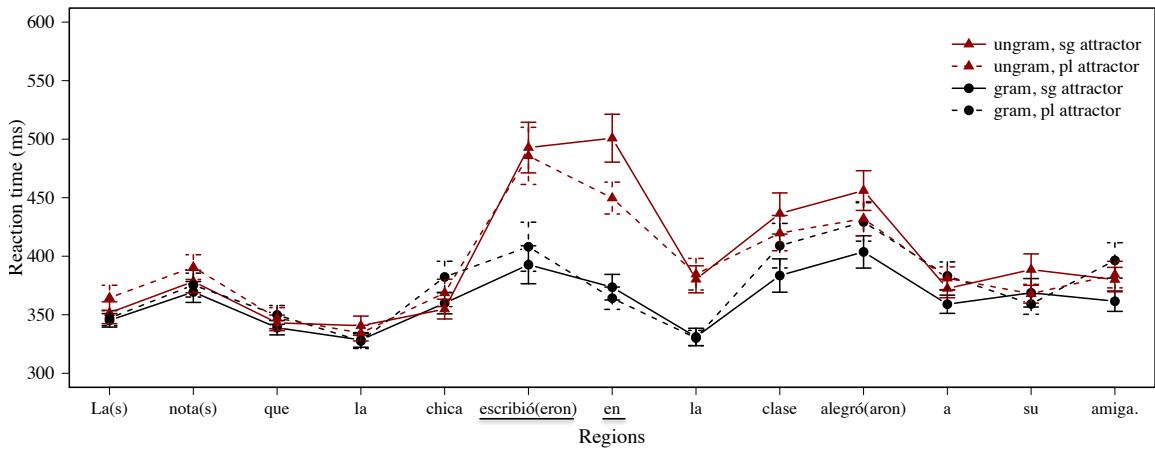


Figure 3 Region by region means in Experiment 1. Error bars indicate the standard error of the mean. Sample sentence: *The note(s) that the girl wrote_{sg/pl} during class cheered her friend up*. The regions of interest are underlined.

	$\hat{\beta}$	CI	SE	<i>t-value</i>
<i>Verb region</i>				
Grammaticality	20	[-7,46]	14	1.45
Attractor number	-4	[-23,15]	10	-0.43
Gram \times Attr number	-16	[-54,22]	20	-0.81
<i>Verb+I region</i>				
Grammaticality	94	[70,117]	12	7.85
Attractor number	-23	[-39,-7]	8	-2.89
Gram \times Attr number	-34	[-64,-4]	15	-2.24

Table 2 Linear mixed-effects model results for the regions of interest in Experiment 1. Significant effects are bolded.

We constructed a vincentile plot to examine the reaction time distributions in the post-verbal region, where both grammaticality and attraction effects were observed. Vincentile plots allow visualization of how a reaction time distribution changes across conditions. For example, if a difference between two condition means is mainly due to a change that affects most of the trials, then the vincentiles corresponding to the two conditions should be parallel to each other across all vincentiles, with the curve corresponding to the slower condition appearing above the curve representing the faster condition. If, on the other hand, a difference between two conditions is due to a change that affects only the slower trials (i.e., the trials with the highest RTs), then the vertical distance between the vincentiles should be small or nonexistent for the vincentiles on the left-hand side of the graph, but larger on the right. Finally, a difference that is due to a change that affects most trials, but has a disproportionate effect on the slower trials, should be manifested in a vertical separation that is present across the full range of vincentiles but that is larger on the right than on the left. The vincentile plot is displayed in Fig. 2 and the results from the statistical analyses are reported in the text.

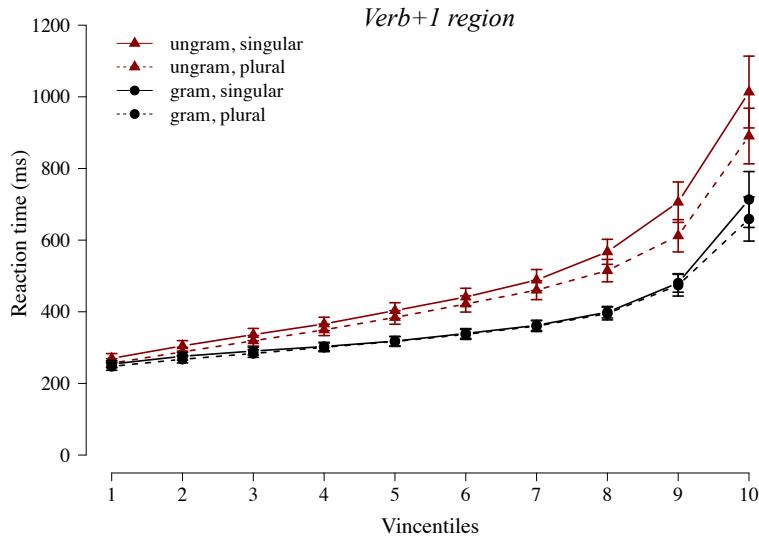


Figure 4 Vincentile plots for the four experimental conditions in the verb+1 region in Experiment 1. Error bars indicate the standard error of the mean across participants.

We compared the effect of grammaticality with the effect of attraction in the grammatical and ungrammatical conditions. The *grammaticality contrast* was operationalized as the difference between the ungrammatical and grammatical curves in the no attractor conditions. The effect of attraction was estimated as the difference between the singular and plural noun conditions in the grammatical and ungrammatical sentences separately (*attractor grammatical contrast* and *attractor ungrammatical contrast* respectively). Visual inspection of the plot suggested that both the effect of grammaticality and the effect of attraction in the ungrammatical conditions affected the late vincentiles more strongly, consistent with a process that impacted mostly the subset of slower RTs. In order to explore the reliability of this pattern we conducted an ANOVA with Vincentile number as a factor (cf. Balota, Yap, Cortese, & Watson, 2008). The results from this analysis confirmed the visual

pattern, as shown by a significant vincentile number \times grammaticality interaction for the *grammaticality contrast* ($\hat{\beta} = 29$, CI = [19,39], $t = 5.78$), and a significant vincentile number \times attractor number interaction in the *attractor ungrammatical contrast* ($\hat{\beta} = -11$, CI = [-20,0], $t = -2.04$). The grammatical conditions, on the other hand, were closely aligned and no interaction was found, consistent with the lack of a significant effect in the overall means.

In addition, our second goal was to address whether detection of an agreement violation preceded the effect of attraction, as predicted by an error-driven retrieval account. To assess whether sensitivity to grammatical violations occurred before attraction, for each contrast we compared the difference score in each vincentile with the difference score in the preceding vincentiles. The divergence point was defined as the earliest vincentile where the comparison reached significance. For the *grammaticality contrast*, the difference was marginal at the seventh vincentile and fully significant from the eighth vincentile onwards ($\hat{\beta}_{V8} = 102$, CI = [36,168], $t = 2.04$). As expected, the attraction effect in the ungrammatical conditions had a later onset, being marginal in the ninth vincentile and only becoming fully significant in the last vincentile ($\hat{\beta}_{V10} = 92$, CI = [17,168], $t = 2.39$). In the grammatical conditions, there was a difference in the last vincentile ($\hat{\beta}_{V10} = 50$, CI = [6,94], $t = 2.23$)².

2.3.3 Discussion

² This effect was unexpected, as no difference was observed in the grammatical conditions in the mean RTs analysis. However, in contrast with the other comparisons, this difference was not replicated in the complementary analysis using only 4 vincentiles, so it is likely to have been driven by a few outlier points in the 10-vincentile analysis.

This experiment shows clear attraction effects in Spanish comprehension and replicates previous findings in English by WL&P. Consistent with the predictions of a retrieval account, we replicated the grammatical asymmetry and found that a plural attractor affected grammatical and ungrammatical sentences differently, as evidenced by the attractor number \times grammaticality interaction. This interaction was due to the fact that a plural attractor facilitated processing in ungrammatical sentences only, having no effect in grammatical sentences. In the ungrammatical cases, the effect of a plural attractor was to decrease RTs immediately after the verb was encountered. This suggests that participants experienced less processing cost when there was a syntactically unlicensed but number matching attractor in the sentence.

The distributional profiles provided by the vincentiles enriched the conclusions from the analysis of the mean RTs. First, the vincentiles replicated the profile observed in the mean reaction time analysis, which found different profiles for the grammatical and ungrammatical conditions. Also, the vincentiles showed that the attraction effect impacted late vincentiles more strongly, as evidenced by the vincentile number \times attractor number interaction in the ungrammatical conditions. This result supports previous findings (Staub, 2009, 2010) that have argued that attraction effects in RCs are driven by a small set of trials, namely, those that disproportionately have longer reaction times.

Lastly, the vincentile plot supported the idea that comprehenders noticed grammatical violations before experiencing attraction effects, consistent with the hypothesis that retrieval is triggered as a repair mechanism after detection of an agreement mismatch. As predicted, the grammaticality effect impacted the vincentiles

earlier than the attraction effect in the ungrammatical conditions. This supports an account where the ungrammaticality of the sentence is detected first, a retrieval is then initiated, and a plural matching attractor noun is sometimes mistakenly retrieved as a result. This timing profile is useful for addressing alternative views such as percolation accounts, which argue that attraction is due to comprehenders' failure to notice agreement errors due to incorrect encoding of the subject phrase. Percolation accounts predict grammaticality and attraction effects to have a similar onset latency, as attraction should act by directly altering the perception of grammaticality. Our results, however, suggest that attraction takes place in those trials in which comprehenders are already experiencing processing difficulty due to the ungrammaticality of the sentence.

Overall, these results are consistent with the hypothesis that a retrieval mechanism underlies number attraction in Spanish. If retrieval is used as a reanalysis process, these results suggest that comprehenders engage in retrieval even when clear and unambiguous morphological cues are present in the input. However, as discussed in the Introduction, we also wanted to examine comprehension profiles when the verbs in the input had a weaker morphological marking, to allow for a closer comparison between English and Spanish. Therefore in our second experiment we used Spanish auxiliary verbs, for which the distinction between plural and singular is less morphologically salient and more closely matched to the previous English experiments.

2.4 Experiment 2: Spanish Auxiliaries

The goal of Experiment 2 was to examine attraction effects elicited by verbs with weaker morphological marking. The materials were almost identical to Experiment 1, with the exception that the main verbs inside the relative clause were replaced by future auxiliary constructions, consisting of an auxiliary followed by the preposition “a”, followed by a non-finite main verb (e.g. “va(n) a escribir”, *is/are going to write*). Therefore, the auxiliary carried agreement information, with singular and plural verb forms being closely matched in length, as in previous English experiments.

2.4.1 Methods

Participants

A different group of participants was tested in this experiment. Participants ($n = 32$, mean age = 27 years, 24 females) were all native speakers of Argentinian Spanish and were recruited from the University of Buenos Aires community. All participants provided informed consent, and were compensated the equivalent of \$5/hour.

Materials and design

The design was identical to Experiment 1, with grammaticality (*grammatical/ungrammatical*) and attractor number (*singular/plural*) as within-subject factors. The same sentence items were used, with the exception that main

verbs were replaced by future auxiliary constructions “va(n) a”. An example set of the experimental materials is presented in Table 1.

Procedure and analysis

The same self-paced reading procedure as in Experiment 1 was used, with one exception. Due to the introduction of the future auxiliary construction “va a”, there was one extra region for analysis: the regions of interest consisted of the auxiliary verb in the relative clause (*verb region*), the preposition “a” (*verb+1 region*) and the non-finite main verb following it (*verb+2 region*). Also, one participant was excluded from the vincentile analysis as he had less than 10 trials in one of the experimental conditions.

2.4.2 Results

In the comprehension questions, mean accuracy was 95.1% and all participants had accuracy rates above 70%. Accuracy did not differ across conditions (all $|z| < 2$).

The region-by-region reaction time averages are presented in Figure 3. The residual mean RTs were analyzed in the three regions of interest. At the verb region no main effects or interactions were observed. At the verb+1 region, there was a main effect of grammaticality, no main effect of attractor number, and an interaction between grammaticality and attractor number. Pairwise comparisons showed that this interaction was due to the number of the attractor having opposite effects in the grammatical and ungrammatical conditions. In the ungrammatical sentences the

plural attractor condition elicited shorter RTs than the singular noun condition ($\hat{\beta} = -16$, CI = [-30,-1], $t = -2.11$). In contrast, in the grammatical sentences the plural attractor condition was associated with longer RTs than the singular noun condition ($\hat{\beta} = 12$, CI = [3,20], $t = 2.73$). In the following verb+2 region, only a main effect of grammaticality was observed. Table 3 shows the results from the mixed-effects model.

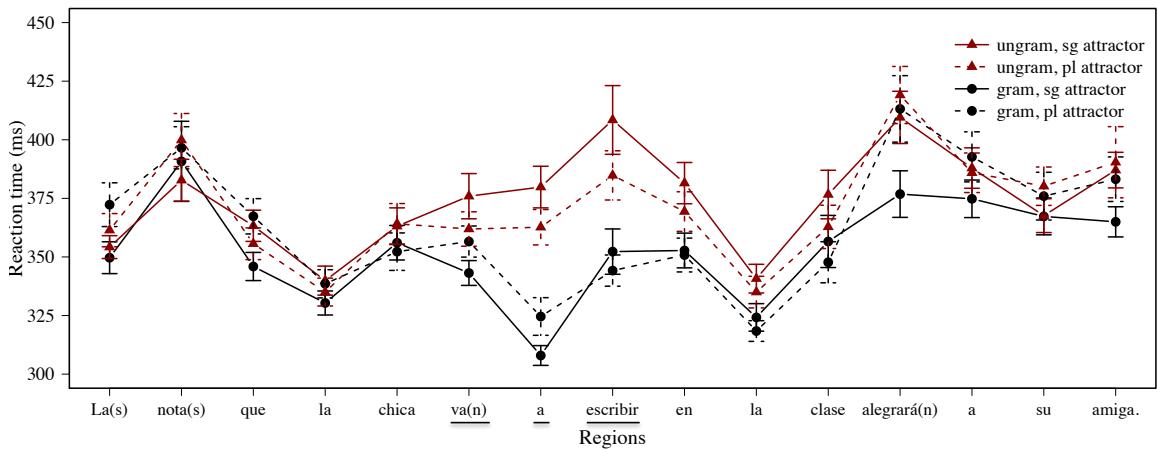


Figure 5 Region by region means in Experiment 2. Error bars indicate the standard error of the mean. Sample sentence: *The note(s) that the girl is/are going to write during class will cheer her friend up.* The regions of interest are underlined.

	$\hat{\beta}$	CI	SE	<i>t-value</i>
<i>Verb region</i>				
Grammaticality	0	[-8,9]	4	0.12
Attractor number	3	[-5,11]	4	0.70
Gram × Attr number	-11	[-27,5]	8	-1.33
<i>Verb+1 region</i>				
Grammaticality	47	[34,60]	7	6.94
Attractor number	-2	[-10,7]	4	-0.44
Gram × Attr number	-27	[-43,11]	8	-3.37
<i>Verb+2 region</i>				
Grammaticality	39	[27,51]	6	6.15
Attractor number	-7	[-18,4]	6	-1.25
Gram × Attr number	-4	[-25,18]	11	-0.34

Table 3 Linear mixed-effects model results for the regions of interest in Experiment 2. Significant effects are bolded.

A vincentile plot was calculated for the verb+1 region, where both grammaticality and attraction effects were observed (Figure 4). The vincentile curves displayed patterns that were consistent with the mean RT analysis, and opposite effects of a plural attractor were observed in grammatical and ungrammatical sentences. Most of the effects seemed to impact the late vincentiles more strongly, consistent with a process that affected mostly a subset of slower RTs. This visual pattern was confirmed by a significant vincentile number \times grammaticality interaction for the *grammaticality contrast* ($\hat{\beta} = 12$, CI = [6,19], $t = 3.60$), and a significant vincentile number \times attractor number interaction for the *attractor grammatical contrast* ($\hat{\beta} = -6$, CI = [-10,0], $t = -2.17$). In the ungrammatical sentences, however, the interaction did not reach significance. Inspection of the plot suggested that although the ungrammatical plural attractor condition was still associated with a downward deflection of the vincentile curve as compared with the singular condition

that was numerically largest in the last vincentile, this difference was smaller than in Experiment 1 and appeared to be more evenly distributed across vincentiles. However, it is also possible that the overall smaller effect size simply reduced our ability to detect an interaction.

Finally, the grammaticality effect showed an earlier onset than the attraction effect in the grammatical conditions. For the *grammaticality contrast*, the difference was significant from the eighth vincentile onwards ($\hat{\beta}_{V8} = 45$, CI = [7,83], $t = 2.31$). For the *attractor grammatical contrast*, the difference was only significant in the last vincentile ($\hat{\beta}_{V10} = -95$, CI = [-133,-56], $t = -4.78$). For the *attractor ungrammatical contrast*, there were no significant differences.

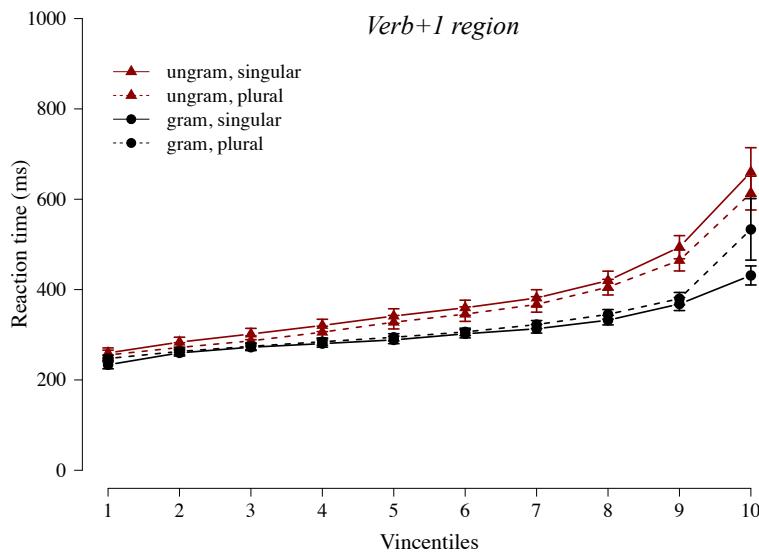


Figure 6 Vincentile plots for the four experimental conditions in the verb+1 region in Experiment 2. Error bars indicate the standard error of the mean across participants.

2.4.3 Discussion

The mean RT analysis in ungrammatical sentences replicates Experiment 1: sentences with a plural attractor elicited faster RTs immediately after the auxiliary was encountered. However, in this experiment there was also an attraction effect in grammatical sentences, as the plural attractor condition elicited slower RTs after the auxiliary verb. This pattern is consistent with increased processing cost in grammatical sentences that contain a plural attractor, a pattern that is not predicted under an error-driven retrieval account.

The distributional profiles provided by the vincentiles showed that plural attractors had opposite effects in grammatical and ungrammatical sentences. The attraction effect in grammatical sentences impacted the vincentiles by increasing the slope of the plural attractor curve. The distribution of this attraction effect was consistent with a process that affected the late vincentiles more strongly, as evidenced by the vincentile number \times attractor number interaction. In ungrammatical sentences, the attraction effect impacted the vincentiles by decreasing the slope of the plural attractor curve; somewhat unexpectedly, in this case we observed no interaction between vincentile and attractor number, although it is possible that the notably smaller attraction effect in this experiment relative to Experiment 1 limited our ability to detect an interaction.

The novel finding in this experiment concerns the grammatical conditions, where longer RTs after the auxiliary suggested more processing difficulty in sentences with plural attractors, a pattern not predicted by an error-driven retrieval mechanism. Under a retrieval account, when information on a verb mismatches the number predicted by the subject noun, a cue-based retrieval is triggered to check

whether the correct feature was somehow missed during the first pass. In the grammatical conditions, however, the verb always agreed in number with the RC subject and therefore retrieval should not have taken place, and no difference in RTs should have been found.

As mentioned in the Introduction, attraction effects in grammatical sentences are predicted under two different views: percolation accounts and the predictive component in WL&P's account. Under a percolation account, comprehenders sometimes wrongly encode the number of the subject phrase when it contains a plural attractor. Therefore, when they read the RC verb, they sometimes mistakenly perceive ungrammatical sentences as grammatical, and grammatical sentences as ungrammatical. However, a percolation account would fail to account for the dissociation observed between main verbs and auxiliaries. Auxiliaries, but not main verbs, showed evidence of processing difficulty in grammatical sentences. Since percolation accounts assume that feature-spreading takes place during processing of the subject phrase (before a verb is encountered) they cannot predict a dissociation contingent on the type of verb. In our experiments, main verbs and auxiliaries had identical sentence preambles, and therefore, similar results should have been obtained in the grammatical conditions.

We propose instead that the differences observed in the grammatical conditions are a result of the predictive mechanism described in WL&P's account, which assumes that in comprehension, subject nouns generate number predictions that are then checked against the verb. In RC configurations, verb number predictions should be generated by both the subject noun within the RC and by the main clause

subject noun (the attractor). When the attractor noun is plural and the RC subject is singular, they will generate predictions for verbs of different number. If the two number predictions can interfere with each other, this revision may sometimes be unsuccessful and encountering a singular verb might (temporarily) violate comprehenders' expectation of a plural verb, increasing reaction times after the auxiliary.

In contrast with the percolation view, this hypothesis can explain the difference between main verbs and auxiliaries if the role of thematic information is taken into account. A crucial distinction between main verbs and auxiliaries is that main verbs carry thematic information about their arguments (e.g., animacy restrictions) whereas auxiliaries do not. Therefore, in a sentence such as "*The notes that the girl writes...*", the thematic information provided by the verb "writes" may have helped comprehenders to quickly revise the plural prediction made by the noun phrase "the notes", enabling them to recategorize it as the object (not the subject) of the verb and to realize that it should not agree with the verb in number. On the other hand, auxiliary verbs like "is/are" do not provide thematic information that could help comprehenders revise the number predictions initiated by the attractor noun.

If interference from plural number predictions generated by the attractor noun is responsible for the processing difficulty in grammatical sentences, an interesting question arises concerning the generality of these predictions in comprehension. Are number predictions equally strong across languages or are they modulated by the morphology of a language? As plural morphology within noun phrases is richer in Spanish and agreement information is functionally more relevant for establishing

subject-verb relationships, there may be a higher payoff for Spanish comprehenders to form robust number expectations prior to encountering a verb. This could result in stronger interference from plural predictions in sentences that have an attractor noun, and an increased difficulty in revising these predictions when the verb does not provide further semantic information. Alternatively, number predictions may carry a similar weight across languages, independent of morphology. Under this account, number predictions in Spanish and English should be equally strong, resulting in prediction errors when comprehenders encounter situations when their predictions initially seem to be violated. This account would predict agreement attraction in grammatical sentences with auxiliaries across languages. However, as WL&P only used main verbs in their English experiments, the existing data do not distinguish these two possibilities. Therefore, in Experiment 3 we examined the effect of number attraction using auxiliaries in English.

2.5 Experiment 3: English Auxiliaries

The goal of Experiment 3 was to investigate the role of morphology in predictive computations, by examining whether English RCs with auxiliary verbs elicit attraction effects in grammatical sentences. If the use of auxiliaries was responsible for the attraction effect in grammatical Spanish RCs, then a similar profile should be found when auxiliary verbs are used in English. Alternatively, if the attraction effect was due to stronger number predictions in Spanish than in English, then we expect to find a grammatical asymmetry in English RCs with auxiliary verbs: attraction should be elicited in ungrammatical sentences but not in grammatical

sentences. To distinguish between these alternatives, we adapted the sentences used by WL&P and replaced main verbs with auxiliary constructions.

2.5.1 Methods

Participants

Participants ($n = 32$, mean age = 21 years, 22 females) were all native speakers of English and were recruited from the University of Maryland community. All participants provided informed consent, and were compensated \$10/hour or received course credit for their participation.

Materials and design

The design was identical to the two previous experiments. We used the same items as WL&P, and we substituted the verbs within the RC with the corresponding past progressive forms (e.g. the main verb “wave(s) to” was replaced by “was/were waving to”). To ensure similarity with Experiment 2, which had a preposition between the auxiliary and the non-finite main verb we inserted an adverb between the auxiliary and the main verb (e.g. “was always waving to”). This was done to keep the same relative distance between the form that carried agreement (the auxiliary) and the form that carried thematic information (the main verb). In order to keep the 48 sentence sets as similar as possible to the original sentences used by WL&P, their items were only replaced when the main verb could not be used in the past progressive form (e.g. “know”). In those cases, the verb and the immediate spillover region consisting of the three words after it were modified. This resulted in partial

modification of half of the item sets. An example set of the experimental materials is presented in Table 4 and the full item set is available in Appendix A.

<i>English auxiliary verb items (Experiment 3)</i>	
Gram, sg attractor	The musician that the reviewer <u>was</u> highly praising last week will probably win a Grammy.
Gram, pl attractor	The musicians that the reviewer <u>was</u> highly praising last week will probably win a Grammy.
Ungram, sg attractor	*The musician that the reviewer <u>were</u> highly praising last week will probably win a Grammy.
Ungram, pl attractor	*The musicians that the reviewer <u>were</u> highly praising last week will probably win a Grammy.

Table 4 Sample set of experimental items in Experiment 3.

Procedure and analysis

The same self-paced reading procedure and analysis as in Experiment 1 were used. However, as in Experiment 2, there were three regions of interest: the auxiliary verb in the relative clause (*verb region*), the adverb following it (*verb+1 region*) and the non-finite main verb (*verb+2 region*).

2.5.2 Results

In the comprehension questions mean accuracy was 94.1% and all participants had accuracy rates above 70%. Accuracy did not differ across conditions (all $|z| < 2$).

The region-by-region reaction time averages are presented in Figure 5. We analyzed the length-regressed RTs in the three regions of interest. At the verb region there were no main effects or interactions. In the following region (verb+1) there was a main effect of grammaticality and a significant interaction between grammaticality and attractor number. The main effect of grammaticality was due to the ungrammatical conditions eliciting higher RTs than the grammatical conditions. The

interaction was driven by the fact that a plural attractor affected RTs only in the ungrammatical conditions ($\hat{\beta} = -37$, CI = [-69,-5], $t = -2.27$). In the grammatical sentences, there was no significant effect of attractor number ($\hat{\beta} = -4$, CI = [-23,16], $t = -0.35$). Finally, in the verb+2 region, main effects of grammaticality and attractor number were obtained. The effect of attractor number was due to the plural attractor conditions having shorter RTs than the singular noun conditions. Table 5 shows the results from the mixed-effects model.

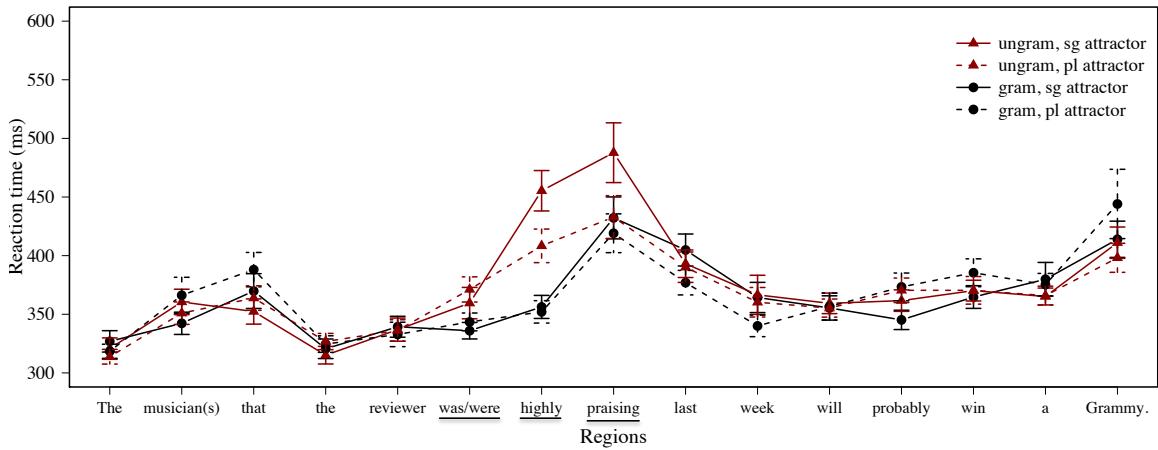


Figure 7 Region by region means in Experiment 3. Error bars indicate the standard error of the mean. The regions of interest are underlined.

	$\hat{\beta}$	CI	SE	<i>t-value</i>
<i>Verb region</i>				
Grammaticality	2	[-9,13]	6	0.33
Attractor number	12	[0,24]	6	1.96
Gram × Attr number	9	[-12,29]	10	0.85
<i>Verb+1 region</i>				
Grammaticality	50	[25,76]	13	3.86
Attractor number	-20	[-42,3]	11	-1.72
Gram × Attr number	-32	[-60,-3]	14	-2.20
<i>Verb+2 region</i>				
Grammaticality	24	[5,42]	9	2.54
Attractor number	-23	[-42,-4]	10	-2.36
Gram × Attr number	-31	[-67,5]	18	-1.68

Table 5 Linear mixed-effects model results for the regions of interest in Experiment 3. Significant effects are bolded.

A vincentile plot was calculated for the verb+1 region, where grammaticality and attraction effects were observed (Figure 6). Inspection of the plot revealed that both grammaticality and attractor number in the ungrammatical sentences had an influence in the vincentile curves. As in the Spanish experiments, these effects seemed to be more pronounced in the late vincentiles, consistent with a process that impacted a subset of slower RTs more strongly. This visual pattern was confirmed by a significant vincentile number \times grammaticality interaction for the *grammaticality contrast* ($\hat{\beta} = 29$, CI = [19,38], $t = 5.95$), and a significant vincentile number \times attractor number interaction for the *attractor ungrammatical contrast* ($\hat{\beta} = -15$, CI = [-26,-5], $t = -2.81$). In the grammatical sentences, no interaction was found.

Lastly, the grammaticality effect showed an earlier onset than the attraction effect. For the *grammaticality contrast*, the difference was significant from the seventh vincentile onwards ($\hat{\beta}_{V7} = 86$, CI = [28,144], $t = 2.92$). For the *attractor*

ungrammatical contrast, the difference was significant from the eight vincentile onwards ($\hat{\beta}_{V8} = 69$, CI = [8,129], $t = 2.23$). For the *attractor grammatical contrast*, there were no significant differences.

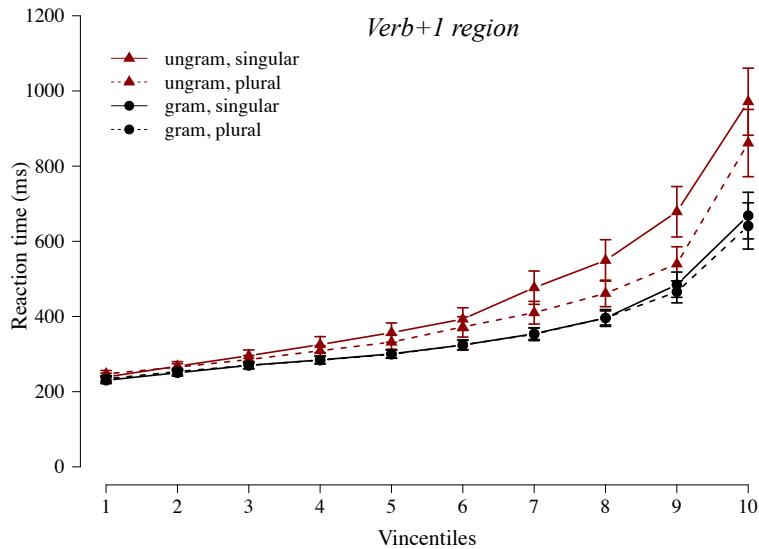


Figure 8 Vincentile plots for the four experimental conditions in verb+1 region in Experiment 3. Error bars indicate the standard error of the mean across participants.

2.5.3 Discussion

As in the previous experiments, a significant attraction effect was observed in ungrammatical sentences, where processing was eased in the presence of a plural attractor. In addition, the vincentile plots revealed that this facilitation affected the late vincentiles more strongly, and it showed a later onset than the grammaticality effect. The crucial result in this experiment, however, is the absence of attraction in grammatical sentences. Neither the mean reaction times nor the vincentile plots showed evidence of processing difficulty in grammatical sentences that contained plural attractors. Importantly, this absence is unlikely to result from insufficient

experimental power, as the lack of an effect in grammatical sentences was obtained in the presence of a robust attraction effect in ungrammatical sentences. These results show that in contrast with Spanish comprehenders, English comprehenders do not experience processing difficulty in grammatical sentences with plural attractors. Under the hypothesis that the processing difficulty found with Spanish auxiliaries was due to prediction errors, these results suggest that number predictions are stronger in Spanish than in English. We discuss the implications of these findings further in the General Discussion.

2.6 General Discussion

The current experiments investigated the question of how morphological variation across languages impacts language comprehension mechanisms by comparing the processing of subject-verb agreement in Spanish and English. We considered two key operations in comprehension, retrieval and prediction, and proposed several ways in which a richer morphological system could affect these mechanisms. Our results demonstrate that comprehenders of a morphologically rich language like Spanish are susceptible to the same attraction errors as comprehenders of English, in that they show facilitated processing of ungrammatical sentences that contain a syntactically unlicensed but number-matching attractor noun.

Our comparison revealed cross-linguistic similarities and differences, which distinguish between the processing of grammatical and ungrammatical agreement. In their responses to ungrammatical agreement, Spanish and English comprehenders behaved similarly across experiments, always showing facilitation in singular-subject

plural-verb mismatch configurations when a plural attractor noun was present. Vincentile plots indicated similar time courses for attraction effects across languages, and these time courses supported the hypothesis that retrieval is engaged as a repair process after an agreement violation is detected in the input. With respect to grammatical agreement, both Spanish and English comprehenders demonstrate no attraction effects when agreement was instantiated on main verbs. Interestingly, however, Spanish but not English comprehenders showed evidence of processing cost when agreement was on auxiliary verbs and a plural attractor noun was present. In other words, while English and Spanish speakers showed a grammatical asymmetry with main verbs, only Spanish speakers exhibited a more symmetric attraction profile with auxiliaries. We suggested that the cross-linguistic similarity in the response to ungrammatical agreement supported the hypothesis that retrieval was engaged uniformly across languages. In contrast, the difference between Spanish and English comprehenders in their response to grammatical agreement was taken to reflect a contrast in the strength of number predictions across languages. We discuss these claims below.

Cross-linguistic similarities

We adopted an account under which comprehenders experience attraction effects when they erroneously retrieve a number-matching attractor noun from memory. The comparison between Spanish and English provides insight into how a richer and functionally more relevant morphology can affect the retrieval mechanism. We predicted that smaller or non-existent attraction effects would be observed if

Spanish comprehenders were more certain about their encoding of number information prior to the verb and they engaged less in retrieval as a reanalysis mechanism. Alternatively, higher attraction effects were expected in Spanish if the morphological cues of the verb were weighted more strongly at retrieval and resulted in more attraction errors. We contrasted these possibilities with the hypothesis that, irrespective of surface morphological differences, the retrieval mechanism was engaged uniformly across languages, which predicted quantitatively similar attraction effects across Spanish and English. Our results rule out a strong version of the first hypothesis, as Spanish comprehenders always showed robust attraction effects in ungrammatical sentences. However, a full comparison of these three hypotheses should also address whether the size of attraction effects was similar across languages.

Although cross-linguistic comparisons are challenging because they involve comparing across experiments with different participants and items, the plot presented in Figure 7 provides a starting point. Here we compare the effects of grammaticality and attraction in Spanish and English across main verbs and auxiliaries. The effect of grammaticality reflects the amount of slowdown due to ungrammaticality, and it was calculated as the mean RT difference between grammatical and ungrammatical sentences in the no attractor conditions. The effect of attraction shows the amount of facilitation in the ungrammatical conditions due to the presence of a plural attractor, and it was calculated as the mean RT difference between the singular and plural noun conditions in ungrammatical sentences. Mean effect sizes were computed using a standardized scale, Cohen's d , where the

difference between the means in the two conditions of interest is divided by the pooled standard deviation (Cohen, 1969; Kirbi & Gerlanc, 2013). 95% CIs were computed by bootstrapping.

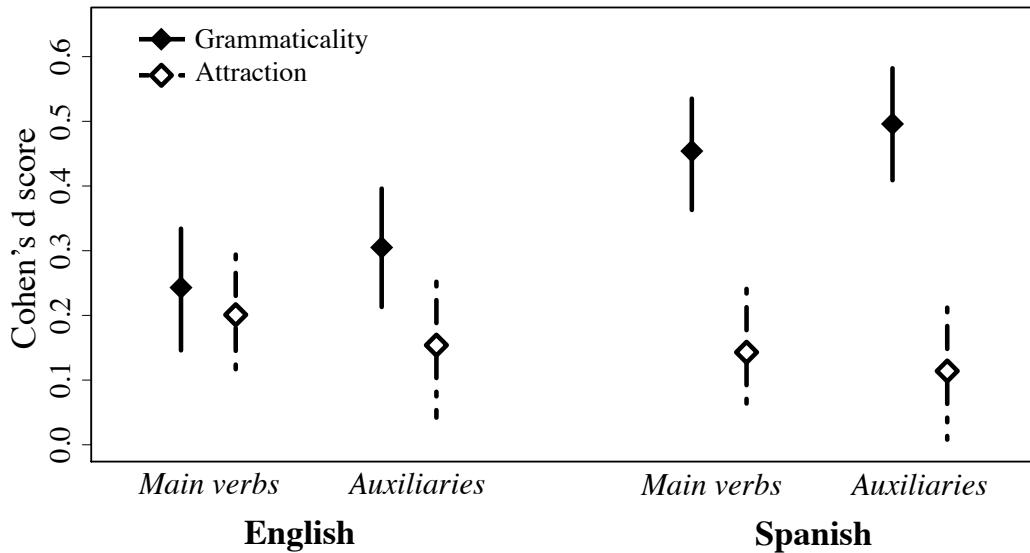


Figure 9 Comparison of grammaticality and attraction effect sizes across four experiments using a standardized scale (Cohen's d). The effect of grammaticality was calculated as the mean RT difference between grammatical and ungrammatical sentences in the no attractor conditions. The effect of attraction was calculated as the mean RT difference between the singular and plural noun conditions in ungrammatical sentences. Diamonds show mean effect sizes, and vertical bars display bootstrapped 95% CIs. Effect sizes were always estimated from length-regressed RTs in the post-verbal region. Data for English main verbs was obtained from WL&P (2009: Experiment 2).

Some useful generalizations about the role of morphology can be drawn from Figure 7. First, a within-language comparison of grammaticality and attraction effects reveals that while in English these effects are closer to parity, in Spanish grammaticality effects are always substantially larger than attraction effects. Second, the comparison of effect sizes across languages reveals a specific way in which morphology impacts agreement computations: although the size of grammaticality effects is always larger in Spanish than in English, attraction effects across languages

are quite similar. Together, this pattern suggests that the effect of a richer morphology is to make comprehenders more sensitive to agreement violations overall, but that the amount of facilitation that comprehenders experience due to attraction is quantitatively similar across languages. These data support the hypothesis that cue-based retrieval is used uniformly across languages in agreement computations, and additionally suggest that retrieval is automatically engaged when comprehenders detect an agreement violation, independent of the morphological markedness of this violation.

Lastly, our results in Figure 7 are consistent with cross-linguistic findings on the role of morphological richness in the production of agreement errors. These results have been mixed (see Lorimor, Bock, Zalkind, & Sheyman, 2008, for discussion), partly due to lexical differences across experimental materials, and also to the existence of linguistic features other than number (e.g. gender, case) that introduce further differences across studies. However, in a recent study, Bock and colleagues (Bock, Carreiras, & Meseguer, 2012) conducted a more controlled comparison between Spanish and English production. They used materials that were matched in literal meaning (Spanish stimuli were created by translating English materials) and controlled for notional number, imageability, and sensibility. Since their goal was to compare the effect of notional number across languages, they used prepositional phrase sentences that varied in whether they allowed for a distributive reading (e.g. “The label on the bottles” vs. “The key to the cabinets”). They found that English and Spanish speakers produced a similar number of plural verbs when given singular head-plural attractor preambles (*Spanish*: 37 responses, 7.4%; *English*:

38 responses, 7.2%) and furthermore, that their likelihood of producing plural verbs was comparably influenced by the distributivity of the sentences, with distributive-nondistributive differences for plural responses in English and Spanish of .09 and .06, respectively.

In contrast with production, where the rate of attraction errors can be straightforwardly quantified as the number of incorrect utterances out of the total number of responses, in comprehension it is more challenging to design an appropriate metric to quantify attraction rates. The current study provides a possible method, by expressing attraction rates as the degree of facilitation in reading times of ungrammatical sentences due to the presence of a plural attractor. Using this method, we observed that, as found by Bock and colleagues, attraction errors are quantitatively similar across Spanish and English. Our results support the claim that while morphological richness makes comprehenders initially more sensitive to agreement violations, facilitation due to attraction errors is surprisingly comparable for English and Spanish speakers. Further comprehension studies will be necessary to address whether this generalization holds across a wider range of languages.

Cross-linguistic differences

The main difference between Spanish and English comprehenders was that Spanish comprehenders showed symmetric attraction effects in grammatical sentences with auxiliary verbs. We considered an explanation based on percolation accounts of agreement (Eberhard, Cutting, & Bock, 2005; Franck, Vigliocco, & Nicol, 2002; Pearlmuter, Garnsey, & Bock, 1999), but we observed that these

accounts would fail to predict any difference between the main verb and auxiliary sentences. As discussed in the Introduction, percolation accounts predict symmetric attraction effects (in both ungrammatical and grammatical sentences) because they assume that errors arise during the encoding of the subject phrase (before the grammatical or ungrammatical verb is encountered). However, it is unclear how these accounts could explain the fact that attraction was symmetric for auxiliary verb agreement but asymmetric for main verb agreement: if errors were due to a wrong encoding of the subject phrase, similar effects should have been found across Experiments 1 and 2 since their sentence preambles were identical. A second way of accounting for symmetric effects would be to assume that a retrieval process is engaged in both grammatical and ungrammatical sentences when the verb is encountered. In cases where the plural attractor is mistakenly retrieved, it should yield facilitation in ungrammatical sentences, as it would agree with the ungrammatical plural verb, and it should yield processing cost in grammatical sentences, as it would disagree with the grammatical singular verb. However, an unrestricted retrieval account predicts similar onsets for grammaticality and attraction effects, as both should result from the retrieval process. As we argue in the next section, the vincentiles provide evidence against this possibility.

In order to capture the English and Spanish data, we suggested that the processing cost observed with Spanish auxiliaries was due to prediction errors. This proposal was based on WL&P's account, which suggested that the computation of agreement in grammatical sentences was implemented through a predictive mechanism, such that encountering the subject noun generates a prediction for the

number of the verb and this prediction is then checked against the verb. We propose that in RC configurations, verb number predictions should be generated by both the subject noun within the RC and by the main clause subject noun (the attractor). When the attractor noun is plural and the RC subject is singular, they should generate predictions for verbs of different number. If these two number predictions can interfere with each other, revision upon encountering the RC verb may be sometimes unsuccessful. As a result, participants may sometimes process singular RC verbs as errors, resulting in increased reaction times. Although these contrasting predictions did not appear to conflict in English, we suggest that the case is different in Spanish. Due to their richer nominal plural morphology, Spanish comprehenders might have more available input and a higher payoff to make stronger number predictions prior to encountering verbs, in a way that English comprehenders do not. This might encourage Spanish speakers to routinely make stronger plural number predictions and to be more unsuccessful in revising them, resulting in processing cost when these predictions initially seem to be violated.

If stronger number predictions are made in Spanish than in English, then the question remains why Spanish comprehenders showed no evidence of processing cost with main verbs in grammatical sentences. We suggested that this difference was due to the role of thematic information. When thematic information is present (as is the case with main verbs), it makes revision easier, as it allows comprehenders to quickly re-evaluate the two nouns in the sentence, categorizing the attractor as the object of the verb and blocking the computation of agreement between them. With auxiliary

verbs, however, thematic information is absent, which could make the recovery from prediction errors more difficult and their cost more observable in reaction times.

Although this predictive account of the cross-linguistic differences we observed is still tentative, the idea that predictions might be influenced by language specific properties has some support in the existing literature (Martin et al., 2013; Vasishth, Suckow, Lewis, & Kern, 2010). For example, it has previously been suggested that verb predictions might be stronger in verb-final languages like German, as German is a language where verbs appear in final position more often (e.g., in subordinate clauses), forcing German comprehenders to maintain verb predictions more robustly than English comprehenders. Some evidence for this hypothesis comes from a study (Vasishth, Suckow, Lewis, & Kern, 2010) where the processing of double center embedded sentences in English and in German was compared. These structures are difficult to process, and a surprising finding is that English comprehenders perceive grammatical sentences (3a) to be less acceptable than their ungrammatical counterparts (3b), which have a missing middle verb phrase (Frazier, 1985; Gibson & Thomas, 1999):

- (3a) The carpenter who the craftsman that the peasant carried hurt supervised the apprentice.
- (3b) *The carpenter who the craftsman that the peasant carried __ supervised the apprentice.

A common explanation for the English pattern is that the increased acceptability of sentences like (3b) is due to speakers having forgotten the prediction of an upcoming verb phrase, reducing working memory load. To address this

hypothesis, Vasishth and colleagues tested the same structures in German, and they found that German comprehenders displayed greater difficulty with ungrammatical than grammatical sentences, showing the reverse pattern than English comprehenders. The authors concluded that strength of verb predictions was modulated by language-specific properties. More recently, Martin and colleagues (2013) have extended this idea and suggested that the strength of predictions might also differ between monolingual and bilingual speakers. While these results are consistent with our findings, further research will be necessary to address the extent to which morphological differences can influence predictive computations. Understanding agreement not as a unitary phenomenon but in terms of the specific prospective and retrospective processes involved in number computations will increase our understanding of the situations where cross-linguistic variation should play a role.

A two-stage mechanism for attraction effects in comprehension

The vincentile plots provide two useful insights into the mechanism underlying attraction effects in comprehension since they allow for a finer-grained comparison between alternative accounts of agreement attraction. The first is that attraction effects take place in only a minority of trials, namely, those that correspond to the longest reaction times, which are indexed by the late vincentiles. The second is that the detection of agreement violations affects the vincentiles earlier than attraction effects.

The finding that attraction effects only influence a small number of trials supports previous findings by Staub (2009, 2010), who used plural attractors in

relative clauses in a forced-choice paradigm in which participants read sentence preambles and made speeded choices between singular and plural verb forms. By modeling the latency of correct responses using the ex-Gaussian distribution, Staub found that increases in latency due to attraction were carried by only a small number of the reaction times, with little or no slowing on other trials. Although we find a similar pattern using a more naturalistic reading paradigm, a comparison across studies should be made cautiously due to several experimental differences. The first is that the nature of our task did not allow us to split trials into correct and incorrect responses, as participants were not asked to make conscious decisions. Therefore, the profiles we observed are necessarily a mixture of both types of responses and cannot be mapped directly to the previous results, which separated those responses. Second, as all the experimental factors in our experiments affected the late vincentiles more strongly, it is possible that properties of the self-paced reading task may prevent cognitive factors from impacting earlier vincentiles at all. Future self-paced reading studies could address this concern by examining the effect on vincentiles of a factor predicted to affect all responses.

The vincentile plots also revealed a timing contrast: attraction effects generally had a later onset than grammaticality effects on the vincentiles. This suggests that attraction due to a plural noun mainly takes place in trials where comprehenders are already experiencing processing difficulty, which is likely due to encountering an agreement violation. This temporal pattern provides evidence in favor of a two-stage, predict-and-retrieve account of agreement attraction in comprehension, and it argues against percolation accounts and retrieval-only accounts

alike. Under percolation accounts, attraction reflects incorrect encoding of the number of the subject phrase, such that comprehenders sometimes fail to notice the ungrammaticality of the sentence. Percolation accounts do not predict a temporal asymmetry, in fact, they would rather seem to predict simultaneous effects of grammaticality and attraction: upon encountering the verb, comprehenders that have encoded the subject correctly should slow down (due to detection of an agreement violation), while comprehenders that have encoded the subject incorrectly should not experience processing difficulty (as the plural attractor and the plural verb should agree in number). Similarly, a retrieval-only view posits that retrieval is always engaged when a verb is encountered. In that case, both grammaticality and attraction effects should reflect the result of a retrieval process, and their onset in the RT distributions should be simultaneous.

In sum, the vincentile plots suggest that participants first detect the ungrammaticality of the sentence and then experience attraction effects. This is consistent with a two-stage process where retrieval is used as a reanalysis mechanism only after comprehenders detect a mismatch between their number prediction and the verb number. We take these results to indicate that retrieval in subject-verb dependencies is better understood as an error-driven second-stage process, and we suggest that an important question for future research is whether a two-stage model is also a better fit for other dependencies (e.g., negative polarity licensing) that have been previously linked to a retrieval mechanism (Vasishth, Brüssow, Lewis, & Drenhaus, 2008). We note, however, that the distributional profile of Spanish auxiliaries observed here might be an exception to this generalization, as the effect in

the ungrammatical conditions did not show the expected time course differences. More work is needed to address whether the lack of an effect was due to insufficient experimental power or to true differences in distributional profile.

2.7 Conclusion

We have shown that despite clear differences in the richness and functional importance of number morphology in Spanish and English, Spanish comprehenders are susceptible to the same attraction errors in subject-verb agreement than English comprehenders, and that the magnitude of these effects is strikingly similar across languages. Furthermore, the distributional profile of attraction effects in Spanish is consistent with a two-stage account in which attraction errors are due to a cue-based mechanism that is triggered by the initial detection of an agreement violation. However, in contrast with English comprehenders, Spanish comprehenders showed a symmetric attraction profile with auxiliary verbs. We proposed that this profile is due to stronger number predictions in Spanish than in English. Together these data suggest that, in comprehension, differences in morphological richness between languages may impact prediction to a greater extent than retrieval.

3 Coreference and the lack of antecedent frequency effects in English

3.1 Outline

In Chapter 2 we examined a possible case of cross-linguistic variation by exploring whether differences in morphology across languages affect the *strength* of number cues at retrieval. We assumed a similar memory encoding of number across languages and focused on whether the strength of number representations increased in a language where morphology was both richer and functionally more relevant. In contrast, in this chapter we examine whether the *content* of memory representations differs across languages. We identify a different property that is also subject to cross-linguistic variation, syntactic gender, and we explore how this property affects the processing of coreference. We then ask whether these facts indicate that speakers of languages with syntactic gender access a qualitatively different type of antecedent representation than speakers of languages without syntactic gender, like English.

This chapter is structured as follows. In section 3.2, we outline the arguments used to claim that coreference with pronouns in English is established at the level of discourse. We review a syntactic account (Sag & Hankamer, 1984) that proposes that only a *discourse-based* representation of a pronoun's antecedent is necessary during coreference, and we outline the impact of this account on several processing models of pronoun resolution in English. In section 3.3, we examine the properties of pronominal coreference in languages with syntactic gender and we show that, in contrast with English, these properties require access to more than a discourse-based

representation of an antecedent noun. We then review experimental findings that support the processing claim that a *lexical* representation of a pronoun's antecedent must be reaccessed. We argue that if previous findings from languages with syntactic gender are taken together with the English findings, these data suggest that different antecedent representations must be retrieved across languages. In section 3.4 we review the previous experimental result (van Gompel & Majid, 2004) that provided the main evidence against lexical re-access in English. We evaluate several potential concerns about the design of this experiment, and in sections 3.5 and 3.6 we present two eye-tracking experiments in English that were conducted to address these issues. Section 3.7 outlines the implications of these experiments for cross-linguistic variation.

3.2 Antecedent representation in English

Coreference offers a key case for examining the content of the memory representations used in language processing: since referential pronouns carry little information of their own, their interpretation depends on the semantic properties of the antecedent that they corefer with. In natural speech, however, a pronoun and its antecedent can be separated by an unbounded number of words. Most current models of memory assume that speakers can only maintain a limited amount of information in their focus of attention (Cowan, 2001, 2005; McElree, 2001, 2006; Wagers & Phillips, 2014), and that information that cannot be maintained in focus must be retrieved from memory. As a consequence, coreference in real time requires that when an element in a linguistic dependency -such as a coreferential pronoun- is

encountered, speakers will often need to recover some type of representation of its antecedent stored in memory.

What kind of antecedent representation is retrieved from memory? In some cases, especially with intra-sentential coreference, a pronoun's antecedent can be an explicit element in the previous linguistic input. In these cases, the interpretation of the pronoun can be obtained from the antecedent's semantic and syntactic lexical properties. The syntactic constraints that govern the selection of suitable antecedents have been previously described by Principle B of binding theory (Chomsky, 1981). Principle B is a negative constraint that states that an antecedent can bind a pronoun as long as it does not c-command it in the pronoun's local clause. For example, if a speaker is presented with the passage in (1), there is a strong intuition that "him" refers to the policeman and not to the deputy:

- (1) **The policeman** told the girls that the deputy had authorized **him** to take them back to the house. **The girls** just asked for **their** coats and silently walked out.

Principle B captures this intuition, as "the deputy" cannot be the antecedent for "him" because it would bind the pronoun from within its local clause. In the second sentence, in contrast, the sentential subject "the girls" can bind the pronoun "their" because it is outside the local clause of the pronoun, the noun phrase "their coats". After an appropriate antecedent has been identified using syntactic and morphological constraints, speakers should be able to interpret these pronouns through their morphosyntactic properties (gender, number, case and animacy) in

combination with the lexico-semantic properties of the antecedent nouns (i.e. long-term information stored about “policeman” and “girl” in the lexicon).

The simplicity of this purely lexical strategy, though, quickly breaks down. Consider the critical sentence in (2), which is structurally very similar to the one above.

- (2) Daisy nervously told **Gatsby** that Tom had invited **him** for dinner. **Gatsby** just raised his eyebrows understandingly and nodded.

In this case, readers’ interpretation of “him” might not solely rely on the lexical properties of the pronoun’s antecedent. For example, their mental representation of Gatsby in (2) will contain information about Jay Gatsby based on what they have read previously in the book: that Gatsby is Daisy’s lover, that he is on bad terms with Tom and therefore it is not advisable for him to come to dinner, etc. While some of these properties might become lexicalized and eventually stored in long-term memory under the word “Gatsby”, many of them might not. What I will call *conceptual* or *discourse-based* properties are constructed during the reading of the book and derive from the events, objects and actions of the plot. A consideration of these kinds of facts suggest that independently of the constraints used to identify the antecedent of a pronoun, speakers’ knowledge about the content and properties of the antecedent exceeds the information that is stored in the lexicon.

An account that can accommodate the facts above is that pronoun interpretation relies on the retrieval of both lexical and conceptual antecedent information. We call this a *lexical re-access account*. It proposes that while comprehenders obtain lexical information by re-accessing a lexical representation of a

linguistic antecedent from long-term memory, they retrieve conceptual information by accessing a discourse-level representation that is constructed during the course of reading. This conceptual representation encodes information about individuals, their properties and the relations between them that is not available in long-term memory but instead constructed online. According to this account, in (2) comprehenders first link “him” to a linguistic antecedent (“Gatsby”), which functions as a pointer to retrieve both lexical information from long-term memory and conceptual information from the discourse representation where further properties about Gatsby are denoted.

The view that there is a reliable mapping between linguistic antecedents and lexical and conceptual representations has its own problems, however. In particular, there are cases where lexical representations are not linked univocally to conceptual entities (3), they appear to be linked to a lexical entity that would yield an incorrect interpretation (4), or more problematically, there is no lexical representation available to link to the appropriate conceptual representation (5):

- (3) This dog likes to sleep, but that dog loves to play; **he** ...
- (4) After science class, the children chopped off the lizard’s tail. As their teacher predicted, a few days later **it** had grown back again.
- (5) John asked his brother for money in a rude manner. **That/this/it** led to an argument.

In (3) it is not sufficient to simply re-access the single lexical entry *dog*, since that lexical item is associated with two different potential antecedents in the discourse. In (4) it would be wrong to identify the pronoun “it” with the lexical noun phrase “the lizard’s tail” in the previous sentence: the pronoun does not refer to a

chopped-off tail, but instead introduces a new entity in the discourse. Finally, (5) is a case of event anaphora (Asher, 1993; Davidson, 1967; Peterson, 1982), where there is not an explicit element in the previous sentence that can function as the pronoun's antecedent: "it" does not refer to "the money" but rather to the event of asking for money in a rude way. In all these instances, it is not straightforward to explain how lexical and conceptual representations could be linked to each other or to a specific linguistic antecedent in the sentence.

Sag and Hankamer accounted for examples like the ones above by suggesting that pronouns do not rely on a linguistic representation of their antecedent and are interpreted directly by accessing conceptual information in a speaker's discourse model (Sag & Hankamer, 1976, 1984). They proposed a distinction between *surface anaphors* (e.g. VP ellipsis, sluicing, gapping and striping), which refer to a linguistic representation associated with an antecedent, and *deep anaphors* (e.g. personal pronouns, sentential "it", some null complements), which refer to some object in a model of the world constructed by the interpreter of the sentence or discourse. This distinction was made on the basis of several syntactic differences in the distribution of these forms. One of these differences was the fact that only surface anaphors require parallelism in syntactic form between an anaphor and its antecedent. For example, in (6), the antecedent clause is in passive form, and therefore ellipsis is only possible when the target clause is also passive, in contrast with "it", which does not require such parallelism. In addition, while surface anaphors require linguistic antecedents and do not sound right in contexts where nothing has been said (7a), deep anaphors can be interpreted as referring directly to extrasentential elements (7b):

- (6) The children asked to be squirted with the hose, so
 - (a) they were ø / *we did ø.
 - (b) we did **it**.

- (7) Hankamer points gun offstage and fires, and a blood curdling female scream is heard.
 - (a) *I wonder who ø / *I wonder who was ø / *Jorge, you shouldn't have ø.
 - (b) I wonder who **she** was / Jorge, you shouldn't have done **it**.

Although Sag and Hankamer's account was designed to explain formal differences in the syntactic distribution of anaphors, it has strong implications for processing. Their account proposes that referential pronouns, as instances of deep anaphora, are interpreted by reference to some object in comprehenders' discourse model, in contrast to surface anaphors that are interpreted by reference to a linguistic representation associated with an antecedent. This means that comprehenders do not need to re-access lexical information about pronoun's linguistic antecedent during coreference. In fact, Sag & Hankamer's view, which we will call a *discourse-only account*, proposes that they don't: "the interpretation of deep anaphors is not mediated by its relation with an antecedent expression at all; it does not, in particular, involve reference to any representation of an antecedent expression, syntactic or semantic, deep or surface" (Sag & Hankamer, 1984:328).

Most of the psycholinguistic research in English after Sag and Hankamer's proposal supported the idea that the nature of the discourse representation that speakers construct during comprehension affects their processing of sentences containing coreferential pronouns (see Garnham, 2001, for review). For example, the prominence of an antecedent in the discourse has been shown to facilitate reading times during coreference (Almor, 1999; Foraker & McElree, 2007; Garrod &

Sanford, 1994; Gernsbacher, 1989; Gordon, Grosz, & Gilliom, 1993; MacDonald & MacWhinney, 1990). In addition, the manner in which speakers perceive time and space also has an effect in processing: when a pronoun is separated from its antecedent by some linguistic material, comprehenders show eased processing (e.g. shorter reading times) to the sentence containing the pronoun when the preceding material alludes to a shorter shift in time (e.g. “forty minutes” vs. “five hours”, Anderson, Garrod & Sanford, 1983) or space (e.g. adjacent vs. non-adjacent rooms, Dutke, 2003; Rinck & Bower, 1995).

Following Sag and Hankamer, many processing models in psycholinguistics have adopted a discourse-only account of anaphora processing in English. For example, in one recent review on the psycholinguistics of anaphora, Nicol & Swinney (2003) address this topic under the heading “Representational Issues” and conclude that the current evidence shows that “a pronoun triggers reactivation of the semantic representation of an NP, not the form itself”. Lucas, Tanenhaus and Carlson (1990) suggest that pronouns immediately map onto a discourse referent and do not go through a linguistic or lexical representation except in cases where the antecedent is out of discourse focus. In the same spirit, when discussing the different uses of coreferring expressions such as proper names as compared with pronouns, Peter Gordon proposes:

Building on the work of others, we (Gordon & Hendrick, 1997) have argued that a discourse model consists of a set of semantic entities and a series of predication, of which the entities are arguments. With respect to lexical selection, we have argued that the primary purpose of names (and other unreduced referring expressions) is to introduce semantic entities into a model of discourse, whereas the primary purpose of pronouns (and other reduced expressions) is *to refer*

directly to entities that are prominent in the discourse model. (in Levelt, Roelofs, & Meyer, 1999; my emphasis)

Although a discourse-only account is compatible with most of the experimental data on pronoun processing in English, in the next section we argue that languages with syntactic gender provide strong conceptual and empirical motivations for a lexical re-access account. We further point out that when processing data from English and from languages with syntactic gender are put side-by-side, they strongly suggest that different types of antecedent representations are accessed across languages.

3.3 Antecedent representation in languages with syntactic gender

Languages that encode gender syntactically pose a problem for a discourse-only account. This account predicts that the information that comprehenders use to interpret a pronoun is obtained solely from re-accessing the conceptual properties of its antecedent in the discourse. This means that information that is encoded in the antecedent's lexical representation and that has no conceptual correlates should no longer be available to comprehenders during coreference. Pronouns carry features such as animacy (e.g. “him” vs. “it”), number (“him” vs. “them”) and gender (“him” vs. “her”) with which antecedents must agree, and abundant psycholinguistic evidence suggests that comprehenders use these features to select appropriate antecedents and rule out feature-mismatching ones (Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Erlich, 1980; Ehrlich & Rayner, 1983; Garrod &

Sanford, 1982; MacDonald & MacWhinney, 1990; Nicol, 1988). However, in some languages, some of these lexical features do not have any clear conceptual correlates.

In English, animacy, number and gender features have conceptual correlates, and thus it is reasonable to assume that they can be represented at the discourse level. For example, plural number in a noun usually correlates with the numerosity of the intended referent³, and gender in English is either semantic (e.g. “boy”, “king”) or stereotypical (“janitor”, “miner”). In languages with syntactic gender, on the other hand, gender is an arbitrary property of nouns and often lacks conceptual correlates. This is most evident with inanimate nouns: in Spanish, for example, the gender difference between table (which is feminine) and pencil (masculine) is marked morphologically on the nouns and their modifiers in a sentence. However, the gender of inanimate nouns has no conceptual consequences, and therefore, there is no motivation for including gender information at the discourse level. Another often-given example is the word “Mädchen” (‘girl’) in German, which denotes female entities but has neuter, rather than feminine, syntactic gender. The arbitrariness of syntactic gender in these languages and its independence from conceptual gender suggest that gender information about an antecedent noun cannot be stored at the level of discourse and needs to be represented in the antecedent’s linguistic form in the sentence or in its lexical entry in long-term memory.

These facts thus raise a critical question for the discourse-only account: how can speakers of these languages establish gender agreement during coreference if they

³ Exceptions to this are nouns that are grammatically plural but conceptually singular (e.g. “trousers”, “scissors”), and nouns that are grammatically singular but are used to refer to multiple objects (e.g. “news”). These cases, however, are limited and constitute the exception, not the rule, in English. In most cases, there is a straightforward alignment between morphological and conceptual number.

can only access conceptual information about a pronoun's antecedent? Consider the Spanish sentence in (8).

- (8) Anabel volvió del supermercado y dejó **el pollo** sobre la mesa y **la carne** en la heladera. A la noche, su novio **lo** usó para cocinar la cena.

Annabel came back from the supermarket and left the chicken_{masc} on the table and the meat_{fem} in the fridge. In the evening, her boyfriend used it_{masc} to make dinner.

Any speaker of Spanish interprets the pronoun “it” in (8) as referring to chicken, not meat. This interpretation can only be obtained from the agreement in gender between “it” and “chicken”, since both chicken and meat are otherwise equally plausible ingredients for dinner. As the gender of these two inanimate antecedent nouns has no conceptual consequences and it is unlikely to be represented in comprehenders' discourse model, it can only be obtained from their representation of the previous sentence or the representations of “chicken” and “meat” in the lexicon. This example thus illustrates a case where conceptual information might be insufficient to license coreference. A stronger illustration involves examples where, even if gender information for inanimate nouns were represented in the discourse, speakers' judgements suggest that gender agreement is established at a linguistic, not discourse level:

- (9) Anabel no sabía si comprar **lomo** o **paleta** en el supermercado. Al final eligió la carne más cara y decidió que **la** / ***lo** iba a cocinar para la cena.

*Annabel didn't know whether to buy sirloin_{masc} or shoulder_{fem} at the supermarket. She finally chose the most expensive meat_{fem} and decided to cook it_{fem} / *it_{masc} for dinner.*

Spanish speakers judge (9) to be acceptable only when the pronoun “it” has feminine gender. This fact is puzzling if gender is licensed at the discourse level:

upon encountering the noun anaphor “the most expensive meat”, comprehenders should be able to reactivate the appropriate (masculine) referent, “sirloin”, in their discourse model, and license the masculine pronoun. In fact, previous evidence suggests that this type of inferences, based on real-world knowledge, occur quickly and automatically after encountering noun-phrase anaphors (Lucas, Tanenhaus & Carlson, 1990). If gender agreement was established at the level of discourse, it is puzzling that the pronoun in (9) cannot agree with its conceptual antecedent (“sirloin”) and instead has to agree with its linguistic antecedent in the sentence, “the most expensive meat”. Therefore, speakers’ judgments in Spanish (and other languages) suggest that even if the gender of inanimate nouns were represented at the discourse level, there are instances where agreement is established with a linguistic antecedent, not a conceptual one.

Offline judgments about examples like those above thus strongly suggest that speakers of languages with syntactic gender need to access linguistic and/or lexical antecedents in order to interpret referential pronouns. This claim has been consistently supported by processing studies in Dutch, French and Spanish (Cacciari, Carreiras & Barbolini-Cionini, 1997; Carreiras, Garnham & Oakhill, 1993; Frazier, Henstra & Flores d' Arcais, 1996; Garnham, Oakhill, Erlich & Carreiras, 1995; Meyer & Bock, 1999). For example, Garnham and colleagues examined whether pronoun resolution is sensitive to the grammatical gender of the antecedent (Garnham et al., 1995). They used sentences that contained two noun phrases before an object pronoun. Both noun phrases were inanimate to ensure that grammatical gender (masculine or feminine) had no correlates with conceptual gender (male or female).

In several self-paced reading studies in French and Spanish, reading times for clauses containing pronouns were faster when the two noun phrases differed in grammatical gender, suggesting that speakers were able to use gender as a cue to disambiguate between the two possible referents in the sentence. In the domain of language production, Meyer and Bock (1999) found that when participants produced pronouns referring back to antecedents consisting of a noun and a prepositional phrase modifier, they made more agreement errors if the antecedent mismatched in gender with the noun inside the propositional phrase. This suggests that speakers of languages with syntactic gender access information about the syntactic gender of the antecedents in order to select pronouns.

3.4 van Gompel & Majid (2004)

We have now reviewed evidence that in languages with syntactic gender both lexical and conceptual information about an antecedent are necessary. In English, on the other hand, retrieval of lexical information does not seem essential, as most of the agreement features that referential pronouns need to check can be specified in the discourse representation. One means of accounting for these observations is to propose that there is a cross-linguistic contrast in the content of the antecedent representations reaccessed during coreference: while speakers of languages with syntactic gender retrieve both lexical and conceptual antecedent information when a referential pronoun is encountered, speakers of English might only retrieve conceptual information. In other words, the exact strategy used to access antecedent information during coreference could vary between languages, or even between tasks

or situations for speakers of a single language. The question, therefore, is whether English comprehenders access the same type of antecedent representations as speakers of languages with syntactic gender, or whether there is systematic variation across languages that can be explained by positing that different representations are accessed during pronoun resolution.

As was pointed out earlier, however, most of the previous English experiments do not provide evidence *against* lexical-reaccess: they just show that conceptual information affects pronoun resolution as well. In particular, this evidence does not address the strong claim advanced by Sag & Hankamer, that pronouns do not retrieve linguistic antecedents. For example, a time advantage in re-accessing antecedents that are prominent in the discourse only shows that discourse modulates antecedent accessibility, but it does not speak to whether lexical information is also accessed. In other words, most of the existing English data is compatible with the claim that pronouns access both lexical and conceptual information, in contrast with Sag & Hankamer's discourse-only proposal.

The study that most directly addressed the viability of a lexical re-access account in English was conducted by van Gompel and Majid (2004, henceforth vG&M), who examined pronoun resolution in an eye-tracking study. They varied the frequency of the antecedent of a pronoun such that in one experimental condition the antecedent was a high-frequency word, (e.g. “criminal”), while in another condition the antecedent was a low-frequency word (e.g. “arsonist”). An example sentence is given in (10), where vertical bars mark the regions of interest in their study:

- (10) The constables suspected the |criminal / arsonist| of the break-in
They searched |in his| bag for| evidence.

Contrary to the predictions of the lexical re-access account, vG&M found an inverse relationship between antecedent frequency and reading times at the region immediately following the pronoun: when the antecedent was an infrequent word, reading times at the post-pronoun region were shorter than when the antecedent was a high-frequency word. No effect was observed immediately at the anaphor region (e.g. Ehrlich & Rayner, 1983; Garrod, Freudenthal, & Boyle, 1994). This is consistent with previous findings, which have usually found the effects of their experimental manipulations in the regions following a pronoun, given that pronouns are typically read quickly, and are frequently skipped in eye-movement studies.

These results appear to argue against lexical re-access in English. Instead, vG&M proposed that the obtained pattern could be captured under a version of the conceptual account which we will refer to as the *prominence account*: low frequency words are less common, and, therefore, they may be more salient in the discourse. In cognitive terms, it has been proposed that salient entities could be mapped to stronger or more active memory representations (Garnham, Traxler, Oakhill, & Gernsbacher, 1996; Garrod, Freudenthal, & Boyle, 1994; Foraker & McElree, 2007). Therefore, since infrequent words are more prominent in memory, co-referring pronouns might be processed faster due to a reduced processing load (Garnham, 2001; Garrod & Sanford 1994; Nicol & Swinney, 2003).

Although vG&M's experiment provides one data point in favor of the discourse-only account, there are also several potential concerns that might prevent us from drawing this strong conclusion. First, if lexical re-access of a pronoun's antecedent is only one of several stages during coreference, frequency effects could

be relatively small and localized, and vG&M’s study may not have had enough resolution to find evidence for these effects in participants’ reading times. Second, inspection of vG&M’s items suggests that on some occasions, the manipulation of lexical frequency resulted in sentences where the low-frequency antecedent created a less common or less plausible situation given the sentence context (e.g. “The family loved the **lawyer/mogul** loyally and steadfastly. They mourned over his death for many weeks”). This may have yielded a manipulation in which lexical frequency and plausibility were confounded: infrequent/implausible antecedents may have been harder to integrate in the sentence when reaccessed at the pronoun, yielding the prominence effect that vG&M found.

As vG&M’s study is key for evaluating the existence of antecedent lexical re-access in English, we followed up on their experiment in an attempt to improve their design and provide a better evaluation of the lexical re-access account. In Experiment 4, we constructed a larger item set and controlled more carefully for the plausibility of the frequent and infrequent antecedents in the sentences. We also used distributional analyses to provide a more in-depth comparison of the reading time profiles at the antecedent and post-anaphor region. If reading times during eye-tracking result from the combination of different stages of processing (some reflecting the effect of lexical re-access and some reflecting eased accessibility of prominent antecedents) then distributional analyses will be more suited to identify these stages. In Experiment 5, we addressed the concern that lexical frequency might be correlated with the plausibility of the antecedents in the experimental sentences. To achieve this, we used proper names, which can still be described using lexical

frequency but have less conceptual information associated with them relative to common nouns.

3.5 Experiment 4: No frequency effects with common noun antecedents

In this experiment we revisited the eye-tracking results previously obtained by van Gompel and Majid. We hypothesized that their reading time patterns could result from multiple processing stages combined together. Our aim was to establish a more detailed timing profile to examine antecedent access during coreference resolution. Therefore, we used distributional analyses to further decompose reading times to identify timing evidence compatible with the lexical re-access account. Before presenting the results of the experiment, we briefly review this technique.

The goal of distributional analysis is to characterize and quantify the effects of experimental manipulations in terms of changes in the shape of empirical reaction time distributions, which are typically both single-peaked and right-skewed. Previous work has shown that the ex-Gaussian distribution provides a good fit to empirical reaction time data. An ex-Gaussian distribution is a right-skewed probability distribution whose probability density function is the convolution of a normal (Gaussian) and an exponential density function. This is equivalent to saying that reaction times that follow an ex-Gaussian distribution can be decomposed as the sum of one quantity that follows a normal distribution, and a second independent quantity that follows an exponential distribution; the ex-Gaussian distribution can be intuitively understood as a combination of two underlying distributions in this way.

The ex-Gaussian distribution is standardly characterized using three parameters. First, just like a normal distribution it has a location parameter, mu (μ), corresponding to the mean of the underlying normal distribution. Changes in this parameter index changes in the location of the distribution independently of any changes in its shape. Second, it has a scale parameter, sigma (σ), corresponding to the variance of the underlying normal distribution. Finally, it has a skew parameter, tau (τ), corresponding to the scale (inverse rate) parameter of the underlying exponential distribution; increasing the value of this parameter has the effect of increasing the skewness of the distribution, independently of any changes in the location of the distribution attributed to μ or symmetrical changes in the variance of the distribution attributed to σ . The overall mean of a distribution, which is the quantity of interest in standard analyses of psycholinguistic experiments, can be obtained by computing the sum of μ and τ .

Previous work has shown that in lexical decision tasks, reaction time distributions for infrequent words often display a shifted location (μ) and an increased skew (τ) compared to those for high-frequency words (Andrews & Heathcote, 2001; Balota & Spieler, 1999; Plourde & Besner, 1997; Yap & Balota, 2007). These previous findings appear to establish a specific profile for changes in reaction time distributions due to frequency effects. Furthermore, recently Staub and colleagues (Staub, White, Drieghe, Hollway & Rayner, 2010) obtained the same distributional pattern in first fixation reading times in two eye-tracking experiments (Drieghe, Rayner & Pollatsek, 2008; White, 2008). Consistent with the findings in single-word

research, infrequent words displayed both a shift in location (μ) and an increase in skew (τ) relative to their high-frequency counterparts.

With regard to the hypothesis of this paper, distributional analysis is a useful tool because it allows us to establish a more detailed comparison between reading time profiles at the antecedent and post-anaphor sites. vG&M's conclusion was based on the analysis of mean reading times. However, patterns observed in the averages might break down under detailed examination of the full distributions, especially if different underlying processes take place with pronouns, as compared with their antecedents. Therefore, our main goal was to establish what changes in the distribution were driving the pattern previously reported in reading times at the antecedent and at the anaphor and post-anaphor regions. A secondary goal of performing distributional analysis was to determine whether the previously observed frequency profile in lexical decision tasks (frequency driving a shift in both μ and τ) would be obtained in sentences, as reported by Staub and colleagues. Given that few studies that have applied this analysis, it is useful to address whether the previously described distributional profile is consistent, particularly when reading times during eyetracking in sentence comprehension are being examined.

3.5.1 Methods

Participants

Twenty-four individuals (13 females, mean age = 22 years) from the University of Maryland community participated in this experiment. All participants were native speakers of English and had normal or corrected-to-normal vision. Data

collected from one additional participant were excluded due to a high percentage of missing trials in one condition (>30%). All participants gave informed consent and were paid \$10 per hour or received course credit for their participation.

Materials

In order to conduct distributional analysis on our materials we constructed a larger set of items, increasing the number of trials to a number that has been accepted as suitable for obtaining reliable ex-Gaussian parameter estimates using current methods (Heathcote, Brown & Mewhort, 2002; Speckman & Rouder, 2004; Staub et al., 2010). The experimental materials consisted of 80 item sets, of which 34 were adopted from vG&M's study. We excluded two of their 36 items because they contained a female antecedent and a feminine pronoun *her* in the second sentence. We chose to only use masculine antecedents because the form *her* is ambiguous between a possessive and an object pronoun, and this ambiguity could have an effect on the reading times at the anaphor and post-anaphor regions, whereas masculine third person pronouns distinguish between possessive (*his*) and object (*him*) forms. In addition, in order to be able to perform distributional analysis we constructed 46 new item sets. The full set of materials is printed in Appendix C. Table 1 shows one item set. Each item set comprised two conditions (*High* and *Low*), and each item consisted of two sentences.

Antecedent frequency	Sample item
High	The villagers warned the captain of the upcoming storm. They worried abo ut <i>his</i> safety out alone in the open sea.
Low	The villagers warned the mariner of the upcoming storm.

|They worried about *his* safety| out alone| in the open sea.|

Table 6 Sample Materials and delimited regions from Experiment 4.

The anaphor of interest was always the pronoun *his*. Its antecedent was always the direct object of the preceding sentence. All pronouns were unambiguous; they could not refer to any other noun phrase, since the subject of the first sentence was always a plural noun.

We manipulated the frequency of the common noun antecedent (high vs. low frequency) in a within-subjects design. We used the SUBTLEX database (Brysbaert & New, 2009; available from the English Lexicon Project, Balota, Yap, Cortese, Hutchison, Kessler, Loftis, Neely, Nelson, Simpson, & Treiman, 2007) to obtain frequency measures for the antecedents. Table 2 shows the antecedent frequency measures for the items taken from vG&M's study as well as for the newly created materials. The frequency of the high-frequency antecedents was much higher than, and did not overlap with, the frequency of the low-frequency antecedents. The difference in log frequency between the two antecedent-noun pairs (e.g. captain/mariner) always differed by at least one, resulting in an overall difference in log frequency between the two conditions of 1.77 (1.85 for the items from vG&M's study and 1.71 for the newly created items). The length of the high and low frequency nouns was controlled within each item set and never differed by more than one character.

	Low Frequency Mean (range)	High Frequency Mean (range)
Materials adopted from vG&M	1.10 (0.02-3.14)	67.19 (11.73-264.96)

Newly created materials	0.54 (0.02-1.67)	33.14 (4.49-208.27)
-------------------------	------------------	---------------------

Table 7 Lexical frequency of antecedents (per million words).

The 80 item sets were divided into 2 lists, such that each list contained exactly one version of each item and 40 items in each condition. Thus, each participant saw each item and each condition, but never saw more than one version of the same item. This amounted to a total of 40 trials per participant per condition. Each pair of high and low frequency antecedents was used in two different items. The experiment also contained 88 two-sentence filler items of comparable length and complexity. Filler items contained other kinds of referential expressions and anaphors, such as female pronouns.

Procedure

Participants were tested individually, and eye movements were recorded using an EyeLink 1000 eyetracker (SR Research, Mississauga, Ontario, Canada), interfaced with a Dell PC. The sampling rate for recordings was 1000 Hz. Stimuli were displayed on a Dell LCD monitor. Participants were seated 32 inches from the computer screen. At this distance 4.6 characters subtended around 1° of visual arc. The angular resolution of the eyetracker is 10–30 min of arc. Viewing was binocular, but only the right eye was recorded. Sentences were presented in 12 pt. fixed width Courier font. Each sentence was displayed on a single line.

The experiment was implemented using the Eye-Track software (<http://www.psych.umass.edu/eyelab/software/>). A calibration procedure was performed at the beginning of each testing session, and re-calibration was carried out between trials as

needed. Before the experiment began, each participant was instructed to read for comprehension in a normal manner. The participant triggered the onset of each sentence by fixating a box on the left edge of the computer screen. Each participant read three practice items before the experimental items were shown. Every experimental and filler item was followed by a yes/no comprehension question to ensure that participants were attending to the stimuli. The comprehension questions never referred to the referential dependency between the pronoun and its antecedent. The order of experimental and filler items was randomized across participants. The entire experimental session lasted approximately 45 minutes.

Analysis

The initial stages of data analysis were carried out using EyeDoctor (<http://www.psych.umass.edu/eyelab/software/>). Fixations were adjusted vertically only in cases where an entire sequence of fixations fell above or below a line of text (i.e. fixations were never adjusted individually). In addition, trials were omitted from the analyses if a long duration track loss occurred at any time during a trial (e.g., if there was no data for half a line of text or more). This resulted in the exclusion of 2.6% of all trials. Fixations shorter than 80 ms in duration and within one character of the previous or following fixation were incorporated into the neighboring fixation. Remaining fixations of less than 80 ms or more than 800 ms were excluded.

We defined three regions for analysis, following vG&M. The *antecedent region* consisted of the antecedent noun only. The *anaphor region* consisted of the pronoun plus three character spaces to its left. Including fixations to the left of the region of interest is a commonly used procedure for analyzing short regions, because

it is assumed that short words such as pronouns are processed during a fixation close to the left of the word when they are skipped (Ehrlich & Rayner, 1983; Garrod, Freudenthal, & Boyle, 1994). The *post-anaphor region* included the word immediately following the anaphor. If this word was shorter than five characters, the following word was included as well. The division into regions for a sample item is shown in Table 1.

Also following vG&M, we analyzed three eye-tracking measures: first-fixation time, first-pass time and total time. First-fixation time is the duration of the reader's first fixation in a region, provided that the reader has not previously fixated on subsequent text. First-pass time is the sum of all fixations on a critical region before the reader leaves it for the first time, either to the left or to the right. For regions consisting of a single word, this measure is identical to gaze duration. Total time is the sum of all fixations in a region. Skips of a region in a particular measure were treated as missing data points.

The raw reading times were analyzed using R (R Development Core Team, 2011). To assess the effect of frequency using traditional linear models (analysis of changes in means), we used the *lme4* package (Bates, Maechler, & Bolker, 2008) to fit a linear model with a fixed effect of frequency (high or low), and random effects of item and subject, separately for each region and fixation measure (that is, we fit linear mixed effects models; see Pinheiro & Bates, 2000). We then derived 95% credible intervals (CIs) and *p*-values for the effect of condition using Markov Chain Monte Carlo (MCMC).

To assess the effect of frequency using distributional analysis, we fit the parameters of an ex-Gaussian distribution by quantile maximum likelihood estimation (Heathcote, Brown, & Mewhort, 2001) for each participant, region, and condition using QMPE (Brown & Heathcote, 2003). Model fits were excluded when they did not converge (as indicated by a numerically singular Hessian matrix; see Brown & Heathcote, 2003). The model parameters from all participants were grouped and tested for effects of frequency in each region.

3.5.2 Results

Comprehension questions

The mean comprehension question accuracy for the experimental items across participants and items was 93.8%, with sentences in the low-frequency condition showing lower accuracy than sentences in the high-frequency condition (high frequency = 95.6%; low frequency = 91.9%; $\beta = -0.53$, $p(\text{Wald}) < .05$).

Reading times

We found an effect of lexical frequency at the antecedent region, and this effect was significant in first-fixation, first-pass and total times, consistently with the results of vG&M. However, there were no effects of lexical frequency in either the anaphor or the post-anaphor region. Therefore, we did not observe the prominence effect that was reported by vG&M. Table 3 shows the mean reading times by condition, and Table 4 shows the statistics calculated using linear mixed effects modeling. To be consistent with our manuscript, we split the experimental items into

those taken from vG&M, the new items we created, and the results from the first 10 items in each condition, to rule out effects of having a large number of trials per condition ($n=40$). All these analyses revealed qualitatively similar results than the analysis with the entire dataset.

	Region	First-fixation duration		First-pass duration		Total time	
		Low	High	Low	High	Low	High
<i>All items</i>	Antecedent	244	229	329	270	542	413
	Pronoun	221	229	247	259	335	338
	Pronoun+1	217	222	252	262	358	362
<i>VG&M items</i>	Antecedent	249	224	339	256	583	405
	Pronoun	217	223	239	254	325	337
	Pronoun+1	216	224	244	257	359	362
<i>New items</i>	Antecedent	241	233	323	280	512	420
	Pronoun	223	233	252	264	342	340
	Pronoun+1	218	221	259	267	358	362
<i>First 10 items</i>	Antecedent	247	228	356	279	631	477
	Pronoun	222	232	254	274	356	376
	Pronoun+1	217	217	258	259	376	409

Table 8 Averages in first fixation duration, first-pass duration, and total time across critical regions in Experiment 4.

	Region	First-fixation duration			First-pass duration			Total time		
		β (ms)	CI (ms)	Pr($\beta < 0$)	β (ms)	CI (ms)	Pr($\beta < 0$)	β (ms)	CI (ms)	Pr($\beta < 0$)
<i>All items</i>	Antecedent	16	[9,23]	<.01	61	[46,76]	<.01	137	[110,163]	<.01
	Pronoun	-9	[-18,0]	0.97	-14	[-28,0]	0.97	-3	[-24,18]	0.61
	Pronoun+1	-6	[-12,1]	0.96	-12	[-26,1]	0.96	-4	[-25,18]	0.65
<i>vG&M items</i>	Antecedent	26	[14,37]	<.01	81	[59,105]	<.01	179	[137,220]	<.01
	Pronoun	-7	[-20,6]	0.86	-17	[-37,5]	0.94	-10	[-41,23]	0.73
	Pronoun+1	-8	[-18,2]	0.94	-12	[-29,4]	0.93	-1	[-34,31]	0.53
<i>New items</i>	Antecedent	9	[-1,18]	<.05	48	[27,69]	<.01	105	[69,138]	<.01
	Pronoun	-10	[-22,4]	0.92	-12	[-31,7]	0.89	3	[-25,30]	0.41
	Pronoun+1	-4	[-13,4]	0.83	-12	[-32,7]	0.89	-7	[-36,22]	0.69
<i>First 10 items</i>	Antecedent	18	[3,33]	<.01	79	[47,113]	<.01	170	[107,220]	<.01
	Pronoun	-11	[-30,6]	0.91	-24	[-52,5]	0.95	-19	[-66,28]	0.80
	Pronoun+1	1	[-12,12]	0.47	3	[-26,29]	0.47	-29	[-79,16]	0.90

Table 9 Statistics for first fixation, first-pass and total time measures in Experiment 4. Reliable effects are in bold font, marginal effects with $.05 < p < .07$ are in bold and italic font.

Distributional analysis

First fixation times were analyzed by fitting an ex-Gaussian distribution for each participant as described above. The resulting parameter values were averaged across participants. At the antecedent region, the effect of low-frequency words on first fixation time was attributed to a 9 ms decrease in μ , a 26 ms increase in τ , and a 7 ms decrease in σ (recall that μ and τ jointly affect the mean of an ex-Gaussian distribution, and τ and σ jointly affect the standard deviation). Only the effect on τ was significant (μ : $t(43) = -1.08, p > .10$; σ : $t(43) = -1.53, p > .10$; τ : $t(43) = 2.67, p < .01$). In contrast, at the anaphor and post-anaphor regions, on the other hand, the distribution of reading times in the two conditions was almost identical, and all parameter differences were non-significant ($p > .10$). Table 5 shows how frequency affects the distribution of first fixation times, as captured by the location, scale, and skewness parameters of the ex-Gaussian distribution (μ , σ , τ). Note that the sample means provided in Table 5 are not exactly the same as the sums of the relevant parameters, due to averaging.

<i>Region</i>	<i>Variable</i>	Mean	μ	σ	τ
Antecedent	Low	244	171	28	76
	High	229	180	35	50
	<i>Frequency effect</i>	15	-9	-7	26
Pronoun	Low	221	162	27	57
	High	229	162	25	58
	<i>Frequency effect</i>	-8	0	2	-1
Pronoun+1	Low	217	166	32	50
	High	222	175	35	37
	<i>Frequency effect</i>	-5	9	3	-13

Table 10 Fitted ex-Gaussian parameter values in Experiment 4, along with the sample means for comparison. Reliable differences are in bold.

Another way of presenting the reading time distributions is by using vincentile plots (Ratcliff, 1979; Vincent, 1912). These are plots of quantiles, estimated in a way that is robust to outliers and intended to support subsequent averaging across participants. Vincentiles are constructed as follows: each participant's data in each condition are divided into the shortest 10% (vincentile 1), next shortest 10% (vincentile 2), etc. The mean of the observations in each vincentile is then computed. Finally, all the participants' means are averaged, for each vincentile, and these values are displayed as connected points on a plot with vincentile index on the *x*-axis and fixation time on the *y*-axis.

Asymmetries in the curve that yield a greater slope at the right side of the graph are indicative of a right-skewed distribution. If the difference between two condition overall means is due to only a shift in location (μ), with no change in skewness (τ), then the curves corresponding to the two conditions should be parallel, with the curve representing the slower condition appearing above the curve representing the faster condition across all vincentiles. If, on the other hand, a difference between two conditions is due only to an increase in positive skew, then the vertical distance between the curves should be small or nonexistent for the vincentiles on the left-hand side of the graph, but the distance should be larger on the right. Finally, a difference that is due both to a change in the location of the distribution and to an increase in right-skewness should be manifested in a vertical separation that is present across the full range of vincentiles but larger on the right than on the left.

The vincentile plots corresponding to all the regions in the items are displayed in Figure 1. Superimposed on the plots are triangles representing predicted vincentile values obtained from the fitted ex-Gaussian distributions. The proximity of these predicted values to the actual vincentile values indicates that the mean ex-Gaussian parameters capture the typical distribution shapes and show that the parameters extracted from the distributions provide a good fit to the data. For all the non-critical regions in our items, the low frequency and high frequency (grey dashed and black solid lines respectively) closely track each other. At the antecedent region, where the model found a change in the skew (τ), there is a marked increase in the slope of the dashed grey line, which is more marked for the right-side vincentiles and has little or no effect on the left-side vincentiles. This corresponds to the low frequency condition having a larger proportion of slower trials. In contrast, there was no difference between the curves at both the anaphor and post-anaphor regions, in agreement with the lack of an effect in the parameters of the ex-Gaussian distribution.

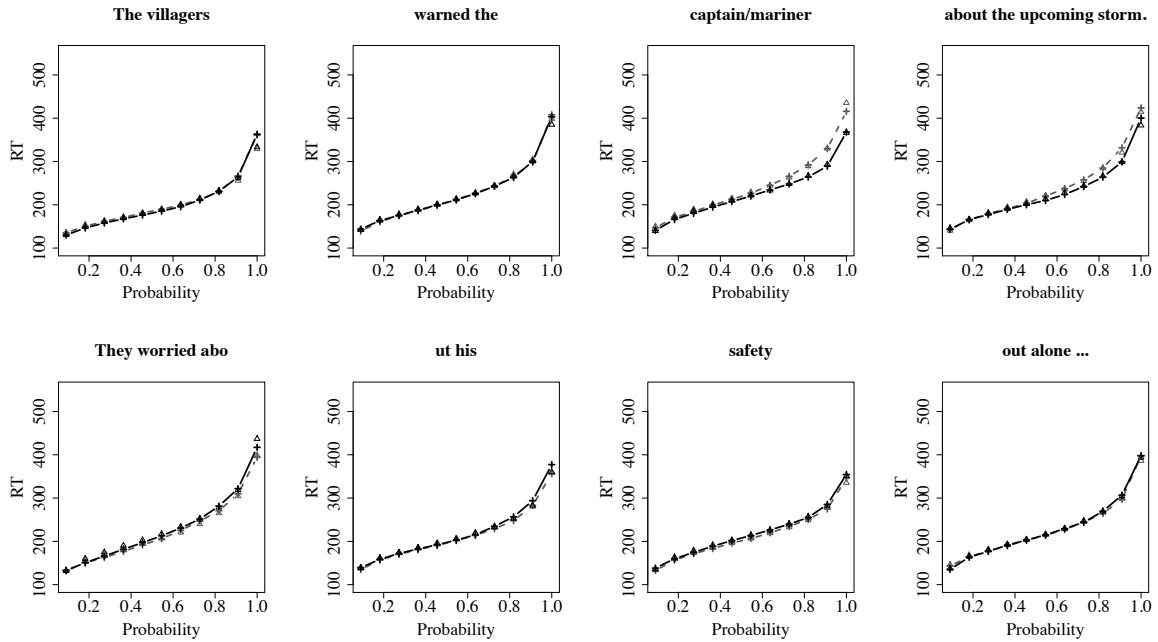


Figure 10 Predicted ex-Gaussian density plots based on parameter values averaged across participants from Experiment 4. The high frequency condition is plotted as a solid line, and the low frequency condition is plotted as a dashed line. The regions of interest are the antecedent (*captain/mariner*), anaphor (*ut his*) and post-anaphor (*safety*) regions.

3.5.3 Discussion

Our results showed a reliable frequency effect at the antecedent noun across various fixation measures (first-fixation, first-pass and total times), but no evidence of a prominence effect at either the pronoun or the immediate post-pronoun region. The distributional analysis revealed that the frequency effect at the antecedent noun was due to a change in the skewness of the distribution in the low frequency condition. Also, consistently with the absence of an effect in the mean reading times, no differences in the ex-Gaussian parameters were obtained at either the pronoun or post-pronoun regions.

While these results do not support a prominence account, they also fail to support the lexical re-access account. To the extent that null effects can be taken as evidence against a specific hypothesis, these results suggest that English comprehenders do not retrieve a lexical representation of a pronoun's antecedent. However, before accepting this conclusion, we attempted to conduct a more controlled manipulation of the lexical frequency of the antecedent noun. Although the frequency of a noun is often characterized as a lexical variable, it also correlates with other non-lexical factors. For example, infrequent nouns tend to depict less common entities in the world. Therefore, Experiment 4 may not have been a pure frequency manipulation.

Experiment 5 addressed this issue by using a similar design to Experiment 4, except that in this case high and low frequency proper names were also used as antecedents. Like common nouns, proper names are also associated with longer retrieval times (Peressotti, Cubelli & Job, 2003), but, unlike common nouns, proper names should be conceptually similar in that they all refer to some individual. When no specific information is known about the individual referred to, then all proper names will be associated by readers with the same concept: ‘some unknown individual.’

3.6 Experiment 5: No frequency effects with proper name antecedents

Experiment 5 aimed to conduct a more controlled manipulation of lexical frequency, in an attempt to dissociate conceptual from lexical frequency. To examine the reliability of the patterns observed in Experiment 4 we also included the previous common noun antecedents. Therefore, we adopted a within-subjects 2 (antecedent type: common noun or proper name) \times 2 (antecedent frequency: low or high) design. No distributional analyses were conducted, since due to the design of the experiment, each participant had only 20 trials per condition, which is less than the number of trials that are recommended to fit the ex-Gaussian distribution.

3.6.1 Methods

Participants

Twenty-four individuals (11 female, mean age = 20 years) from the University of Maryland community participated in this experiment. All participants were native speakers of English and had normal or corrected-to-normal vision. All participants gave informed consent and were paid \$10 per hour or received course credit for their participation.

Materials

The materials were derived from the items used in Experiment 4 to produce a 2×2 design with antecedent type (common noun vs. proper name) and antecedent frequency (high vs. low) as within-subjects factors. Each frequency pair of common nouns used in Experiment 4 was paired with a pair of proper names that did not differ from one another by more than one character in length but contrasted sharply in lexical frequency. A sample item for each antecedent type condition is shown in Table 6.

Measures of the lexical frequency of proper names were estimated using the number of results reported by a Google search within the domain of facebook.com in the United States (see Uyar, 2009 for discussion of the accuracy of approximate search engine counts). We used Facebook instead of the SUBTLEX database because we reasoned that Facebook would better reflect the actual frequency distribution of the proper names that our undergraduate participants were exposed to. Care was taken to avoid the use of proper names strongly associated with prominent historical figures (e.g., Adolf, Judas, Napoleon, etc.). The name counts (number of hits returned in the search) were logged. Frequency estimates for the proper names in the high frequency

condition (average log frequency = 7.42, range between 7.13 and 8.49) were much higher than, and did not overlap with, those in the low frequency condition (average log frequency = 5.66, range between 4.9 and 5.99). The difference in log frequencies between the two conditions was at least 1 for all pairs, resulting in an average difference in log frequency of 1.76.

The 80 item sets were divided into 4 presentation lists, such that each list contained exactly one version of each item. There were 20 items in each condition. Each list also contained 80 two-sentence filler items of comparable length and structural complexity. Every experimental and filler item was followed by a yes/no comprehension question, and the order of items was randomized across participants.

Antecedent type	Sample material
Common noun	The villagers warned the <i>captain/mariner</i> of the upcoming storm. They worried abo ut <i>his</i> safety out alone in the open sea.
Proper name	The villagers warned <i>Martin/Roscoe</i> of the upcoming storm. They worried abo ut <i>his</i> safety out alone in the open sea.

Table 11 A sample item in Experiment 5 with critical regions indicated.

Procedure

The procedure was identical to Experiment 4.

Analysis

Data processing and analysis were the same as in Experiment 4, with two exceptions. First, we included an additional spillover region for analysis, resulting in two post-anaphor regions: *pronoun+1*, consisting of the word immediately following

the pronoun, and *pronoun+2*, consisting of the word immediately following *pronoun+1*. As before, if any of these words was shorter than five characters, then the following word was included as well. Given the length of our sentences, including a second spillover region was not possible in Experiment 4, because it would have resulted in the *pronoun+2* region overlapping with the sentence final region in some items. Therefore, seven of the items from Experiment 4 were lengthened. An example of the division of regions in a sample item is shown in Table 6.

Second, the statistical model was fit to the data separately for each antecedent type (common noun or proper name), with random intercepts for subjects and items, and with antecedent frequency as a fixed effect. Long duration track loss in individual trials resulted in the exclusion of 1.25% of all trials.

3.6.2. Results

Comprehension questions

Participants answered the comprehension questions with mean 93.3% accuracy (*common noun*: high frequency = 94.3%, low frequency = 93.9%; *proper name*: high frequency = 91.1%, low frequency = 94.1%). There were no significant differences between conditions (antecedent type: $\beta = -0.29$, $p(\text{Wald}) > .10$; antecedent frequency: $\beta = 0.28$, $p(\text{Wald}) > .10$).

Reading times

Table 7 shows the mean first fixation, first-pass and total reading times for all four conditions across the critical regions. Table 8 shows the results of the mixed effects model.

Region	First fixation duration		First-pass duration		Total time	
	Low	High	Low	High	Low	High
<i>Common noun</i>						
Antecedent	265	239	345	278	612	459
Pronoun	253	260	281	295	375	389
Pronoun+1	235	247	275	278	382	368
Pronoun+2	260	254	302	305	420	432
<i>Proper name</i>						
Antecedent	260	240	289	257	480	399
Pronoun	254	257	286	290	401	401
Pronoun+1	234	244	265	273	400	400
Pronoun+2	259	249	309	300	449	450

Table 12 Averages in first fixation duration, first-pass duration, and total time across critical regions in Experiment 5.

Region	First fixation duration			First-pass duration			Total time		
	β (ms)	CI (ms)	Pr($\beta < 0$)	β (ms)	CI (ms)	Pr($\beta < 0$)	β (ms)	CI (ms)	Pr($\beta < 0$)
<i>Common noun</i>									
Antecedent	26	[13,38]	<.01	67	[47,88]	<.01	157	[114,200]	<.01
Pronoun	-5	[-21,10]	0.26	-11	[-32,11]	0.15	-11	[-45,22]	0.26
Pronoun+1	-10	[-22,1]	<.05	-1	[-21,18]	0.46	11	[-19,43]	0.24
Pronoun+2	5	[-8,18]	0.22	-4	[-21,14]	0.34	-12	[-36,21]	0.25
<i>Proper name</i>									
Antecedent	21	[8,34]	<.01	33	[16,49]	<.01	87	[49,123]	<.01
Pronoun	-2	[-16,13]	0.40	-3	[-22,18]	0.42	0	[-34,34]	0.49
Pronoun+1	-8	[-19,1]	<.05	-6	[-23,12]	0.25	4	[-34,41]	0.43
Pronoun+2	10	[-3,22]	.06	11	[-9,31]	0.14	-3	[-45,35]	0.43

Table 13 Statistics for first fixation, first-pass and total time measures in Experiment 5. Reliable effects are in bold font, marginal effects with $.05 < p < .07$ are in bold and italic font.

Common noun antecedents

At the antecedent region an effect of frequency was found in first fixation, first pass, and total reading times, due to longer reading times for low frequency antecedents. At the pronoun region, there was no effect of antecedent frequency on any of the three measures. The anaphor region was skipped on 28.4% of trials in the

high frequency condition, and 26.2% of trials in the low frequency condition. At the pronoun+1 region, a prominence effect was observed in first-fixation times, with longer reading times in the high frequency than in the low frequency condition.

Proper name antecedents

The pattern of results in the proper name conditions was similar to that in the common noun conditions. At the antecedent region a significant effect of frequency was found in first fixation, first-pass and total reading times, due to longer reading times for low frequency antecedents. At the pronoun region there was no effect of antecedent frequency on any of the three measures. The anaphor region was skipped on 24.1% of trials in the high frequency condition, and 25.1% of trials in the low frequency condition. At the pronoun+1 region, a prominence effect was observed in first-fixation times, with longer reading times in the high frequency condition than in the low frequency condition. In contrast, first-fixation times in the pronoun+2 region displayed a marginally significant frequency effect, with shorter reading times in the high-frequency condition.

3.6.3. Discussion

The results from Experiment 5 showed a lexical frequency effect at the antecedent, both in the common noun and proper name conditions, and it was again significant in the three measures of interest. Unlike Experiment 4, we now found a prominence effect consistent with vG&M in the post-anaphor region, which was read more quickly when the antecedent was an infrequent word. This effect was only

significant in first-fixation times (in contrast with vG&M) and it was obtained in both the proper noun and common name conditions. In contrast, the only evidence of a lexical frequency effect was observed in the pronoun+2 region in the proper name conditions. This effect, however, was marginal and only present with proper names. In addition, it occurred two regions downstream from the pronoun region, which complicates its interpretation as a correlate of lexical-reaccess triggered by the pronoun. In sum, the main reliable finding of Experiment 5 was a prominence effect consistent with the previous results by vG&M.

3.7 Conclusion

The two experiments presented above along with previous results by vG&M demonstrate that pronouns with infrequent antecedents do not cause processing cost during reading, in that they are not read more slowly than pronouns with frequent antecedents. These results obtain when antecedents are common nouns (Experiment 4) and also when they are proper names (Experiment 5), a manipulation that allowed for a cleaner dissociation between lexical frequency and plausibility. In sum, comprehenders do not show evidence of reaccessing a lexical representation of a pronoun's antecedent during coreference in English. In what follows, we explain how these findings inform the question of whether there is variation in antecedent representation across languages.

The lack of a lexical frequency effect in pronoun processing has different implications for the lexical re-access account depending on the assumptions that are made about the architecture of the lexicon. Here, in contrast with accounts that propose a single lexical level, where all word properties are stored together (e.g., Caramazza, 1997; Caramazza & Miozzo, 1997), we will adopt a model of lexical access that proposes that semantic, syntactic and phonological properties of a word are stored at separate levels of representation and are accessed differently in processing (for review, see Levelt, 2001; Levelt, Roelofs, & Meyer, 1999). Although it is conceived as an account of lexical access for speech production, we adopt it here as it is a model that is widely accepted and psychologically grounded and it makes explicit timing predictions for processing.

According to this model, lexical information in long-term memory is represented in a three-level architecture: a conceptual stratum, a *lemma* stratum and a *lexeme* stratum. Lexical concepts are single (non-decompositional) nodes that express the semantic properties of words. They encode different semantic relationships such as hyper- and hyponymy (e.g. that a canoe is a type of vessel) and synonymy (e.g. that the meaning of the concept ‘canoe’ is related to the meaning of the concept ‘ship’, see Figure 2 below). An active lexical concept spreads some of its activation to its corresponding lemma node, which represents the syntactic properties of a word in a language; for example, the lemma of the word “ship” in Dutch encodes the fact that it is a noun of neuter gender, whereas the lemma of a verb such as “escort” is associated with specific person, tense and number features, as well as the verb’s subcategorization frame. Selection of a specific lemma takes place under competition

with other lemmas that are co-activated by related concepts. Finally, selection of a particular lemma causes retrieval of the phonological properties of a word (e.g. phonemes, lexical stress and syllabic structure), encoded by its lexeme. Therefore, this model describes a word's lexical entry as an item in the mental lexicon that consists of a lexical concept, a lemma, and a lexeme representation. In addition, the architecture is serial in that it assumes that activation spreads from concepts to lemmas, and from lemmas to lexemes. An outline of the model is shown in Figure 2.

A crucial property of this model is that it assumes separate storage of syntactic (lemma) and phonological (lexeme) information. Several types of evidence support this dissociation: patterns in speech errors (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Garrett, 1980), semantic interference effects in the picture-word interference paradigm (Glaser & Düngelhoff, 1984; Roelofs, 1992, 1997), findings on tip-of-the-tongue states (Badecker, Miozzo, & Zanuttini, 1995; Vigliocco, Antonini, & Garret, 1997), gender congruity effects in computing agreement (Jescheniak, 1994; Schriefers, 1993; van Berkum, 1997), specific frequency effects in accessing gender information (Jescheniak & Levelt, 1994), and studies using the lateralized readiness potential that have shown earlier brain responses to gender than to phonological information (van Turennout, Hagoort, & Brown, 1997, 1998). Overall, these studies have found that syntactic and semantic information are dissociable (e.g. both healthy and impaired speakers can have access to one but not the other), that they are affected differentially by variables such as word frequency, and that access to syntactic features temporally precedes access to phonological properties in production.

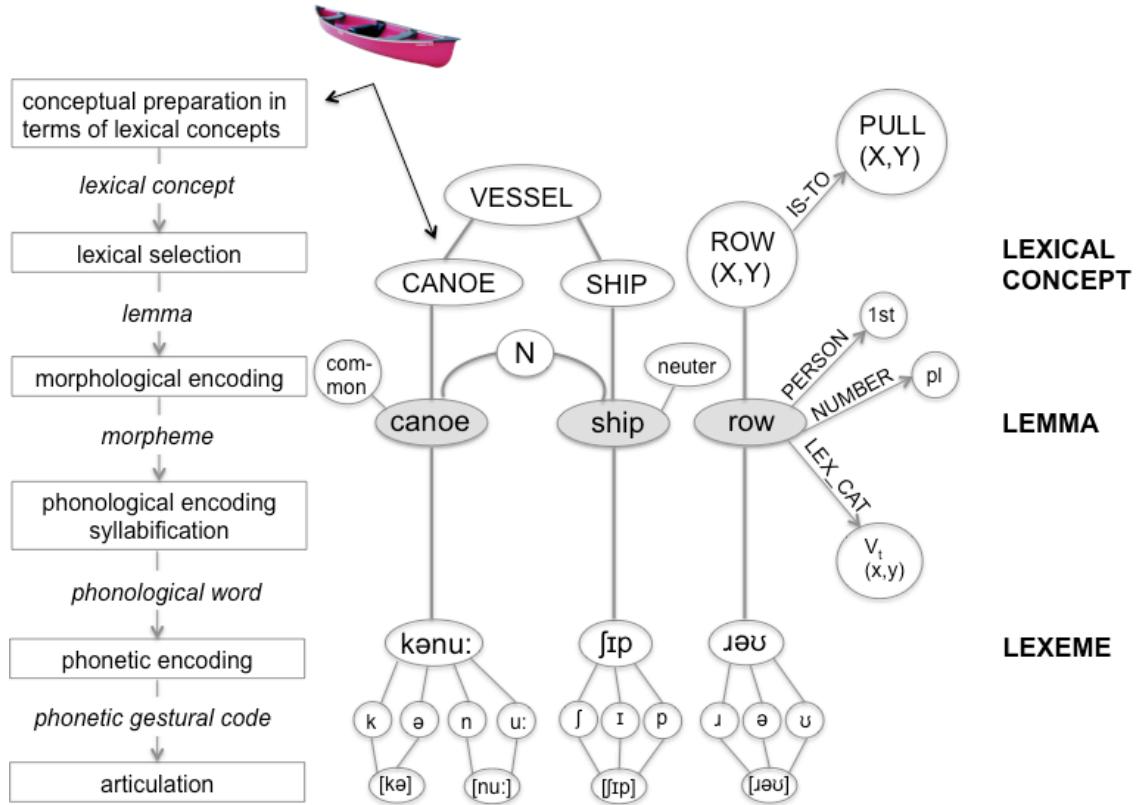


Figure 11 A cognitive architecture for the lexicon. Levelt's model proposes a feed forward activation-spreading network with three levels. Nodes in the top conceptual level represent lexical concepts. Nodes in the lemma level represent syntactic properties of words. Nodes in the lexeme level represent phonological and phonetic properties. Left panel: processing stages in the network. Right panel: example fragment of lexical network. Lemmas are shaded in grey and gender information corresponds to the Dutch equivalents of the English words. Adapted from Levelt (1998: 171).

What does the lemma-lexeme dissociation imply for pronoun processing?

Crucially, it raises the possibility that even if pronoun resolution involves lexical re-access, not all types of antecedent lexical information need to be reaccessed during coreference. In particular, it is possible that antecedent reactivation involves retrieval of semantic and syntactic, but not phonological features. This is because Levelt's model considers that word frequency is a property of the lexeme, not the lemma node

(Jescheniak & Levelt, 1994; Levelt, Roelofs, & Meyer, 1999). Therefore, under Levelt's model the lack of antecedent lexical frequency effects is compatible with the idea that the antecedent's lemma, but not its lexeme, is retrieved during coreference. Lexical re-access could take place during pronoun resolution in English, but where lexical retrieval would selectively target syntactic and semantic information while ignoring phonological properties. In the next section, we elaborate on the implications of the lemma-lexeme distinction for issue of cross-linguistic variation.

A last point is in order to clarify a terminology issue. Throughout this chapter, we have used the terms “conceptual” and “discourse-based” interchangeably. However, the conceptual level in Levelt's model should not be confused with these terms. Levelt's conceptual level is still lexical, not discursive, in that it describes semantic information about words that is context independent and stored in long-term memory. In contrast, as was outlined in section 3.2, a discourse-based antecedent representation refers to the representation of a pronoun's referent that results from the integration of the entities and actions conveyed by the linguistic context of a sentence. This representation, in contrast with Levelt's lexical concepts, is not a property of a specific word, and it results from the on-the-fly computation of sentential information, not the retrieval of semantic features from long-term memory. Although there are clear connections between lexical and discursive knowledge, as both are related to the computation of meaning, there is a distinction between them in that not all lexical-conceptual features of a word are incorporated to a speaker's discourse representation (e.g., Cotter, 1984; Garrod & Sanford, 1982, 2000; McKoon & Ratcliff, 1981) and conversely, not all properties about a pronoun's antecedent that

are constructed at the discourse level should be stored in a comprehenders' lexicon after the reading of a text (cfr. the Gatsby example in section 3.2). Unless explicitly noted, in the following sections the words "conceptual" and "discourse-based" are used as synonyms, and the information in the lexical concepts in Levelt's model is simply described as "semantic" information to avoid ambiguity.

Possible scenarios for cross-linguistic variation in antecedent representation

Given the model of lexical access outlined above, there are several possible scenarios according to which the same or different types of antecedent representation is reaccessed across languages. Importantly, while some of the scenarios are ruled out by our previous experiments and vG&M's findings, others are still viable and require more evidence to be evaluated. Figure 3 presents a summary of these scenarios, which we outline below in turn.

First, the absence of an antecedent lexical frequency effect rules out the possibility of full lexical re-access (retrieval of semantic, syntactic and phonological information) in English. As Levelt's model locates word frequency at the lexeme level, the absence of a frequency effect means that lexeme information about an antecedent is not retrieved during pronoun processing in English.

In contrast, in the next section we present cross-modal evidence that the semantic properties of a pronoun's antecedent seem to be reactivated during co-reference in English. There are at least two possible means of accounting for these semantic effects. One possibility is that comprehenders access this information by retrieving the lexical concept and the lemma representations of the pronoun's

antecedent from the lexicon. Alternatively, semantic effects in cross-modal priming studies could be due to comprehenders retrieving a conceptual representation of a pronoun's antecedent from their discourse model. This second alternative would require semantic features of the antecedent to be specified redundantly in the lexicon and in the discourse representation that comprehenders construct online. Currently, there is no evidence that speaks to this issue, so none of these alternatives can be ruled out.

While the English data indicates that lexeme information is not retrieved, it is not yet clear from the data presented so far whether lexeme information is retrieved in languages with syntactic gender. In section 3.3 we presented arguments that such languages at least require retrieval of the antecedent's lemma, to capture the facts about gender agreement between pronouns and their antecedents. However, no comprehension studies to date have addressed whether the antecedent's lexeme is retrieved as well. If both the lemma and the lexeme were retrieved, then languages with syntactic gender would involve full lexical re-access, while English comprehenders would involve either discourse-only or discourse and lemma re-access. On the other hand, comprehenders of languages with syntactic gender might only re-access the antecedent's lemma during co-reference. If English speakers are shown to also have access to the antecedent's lemma, then we could maintain an account according to which the same type of antecedent representation is retrieved across languages. We evaluate some of these scenarios in the next chapter.

ANTECEDENT REPRESENTATION ACROSS LANGUAGES

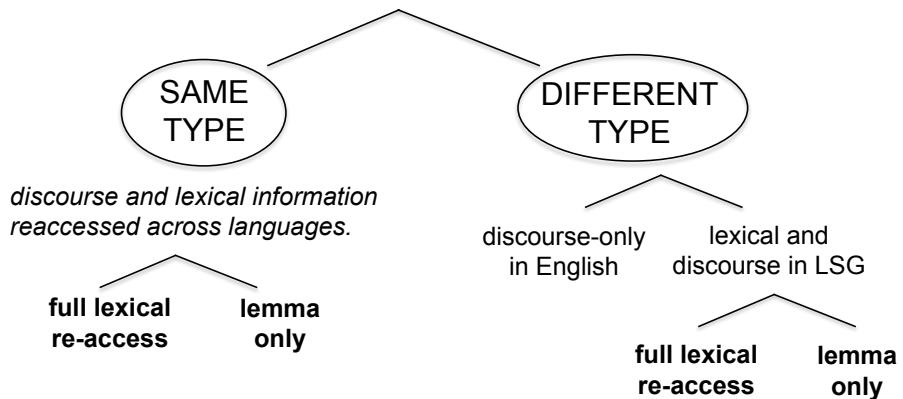


Figure 12 Possible scenarios in antecedent retrieval across languages. LSG stands for languages with syntactic gender.

4 Antecedent representation in a language with syntactic gender

4.1. Outline

This chapter examines whether speakers of a language with syntactic gender retrieve a qualitatively different type of antecedent representation than speakers of English. Previous studies have suggested that speakers of languages with syntactic gender reaccess both lexical and discourse information about a pronoun's antecedent, whereas only discourse information is necessary in English. In two eye-tracking experiments we examine the types of lexical information that speakers of German and English retrieve during reading. We show that in German, readers show a processing difference between semantic and phonological information: whereas comprehenders' eye movements patterns seem immediately sensitive to semantic features, no form effects were observed. In English, comprehenders do not show sensitivity to either semantic or phonological antecedent lexical features, consistently with the claim that they only retrieved a discourse-based antecedent representation. We propose that the processing of pronouns varies cross-linguistically and that antecedent retrieval depends on the type of information relevant to the grammar of each language.

4.2 Introduction

A major challenge in the study of sentence comprehension is to specify the nature of mental representations that speakers create when understanding language. Some representations have stable meanings stored in long-term memory, such as the word “singer” or the verb “told”. Others, however, pose a greater challenge, as their meaning depends on speakers’ dynamic interpretation of contextual material. Pronouns constitute one of these cases, since they carry little information of their own and their interpretation depends on the properties of their antecedents. In natural speech, a pronoun and its antecedent can be separated by an unbounded number of words. Since information quickly decays in memory (Cowan, 2001, 2005; McElree, 2001, 2006; Wagers & Phillips, 2014), coreference in real time requires that when a pronoun is encountered, speakers will often need to recover some type of representation of its antecedent from memory. In this study we ask: what type of antecedent representation is retrieved during comprehension?

We focus on the distinction between lexical and discourse-level representations. Lexical information refers to the features of individual words that are stored in speakers’ long-term memory. In contrast, discourse-level representations are constructed on-the-fly using information provided by the sentence context, and they encode information about individuals, events, their properties and the relations between them. Previous research in English has supported the idea that coreference involves accessing a conceptual representation of a pronoun’s antecedent in a discourse model (the *discourse-only hypothesis*; Cloitre & Bever, 1988; Lucas, Tanenhaus & Carlson, 1990; Sag & Hankamer, 1984). However, there is less

consensus regarding whether a pronoun also reactivates a lexical entry of its antecedent (the *lexical re-access hypothesis*).

The lexical re-access hypothesis has been proposed for languages with syntactic gender, such as German, Spanish and Italian. In these languages gender is an arbitrary property of nouns and often lacks conceptual correlates (e.g. the word for “table” is feminine in Spanish but masculine in German). This fact makes it unlikely that gender information is represented at the discourse level. Since syntactic gender information is stored in a word’s lexical entry, it motivates the need for speakers to retrieve a lexical representation of a pronoun’s antecedent. As a consequence, previous findings that speakers of these languages use gender agreement between pronouns and antecedents as a cue for disambiguation in comprehension (Cacciari, Carreiras & Barbolini-Cionini, 1997; Carreiras, Garnham & Oakhill, 1993; Frazier, Henstra & Flores d’ Arcais, 1996; Garnham, Oakhill, Erlich & Carreiras, 1995) and pronoun selection in production (Meyer & Bock, 1999) have been taken to suggest that morphosyntactic information of a pronoun’s antecedent in the lexicon must be reactivated during coreference in addition to its discourse-based representation.

While these facts support the need for retrieval of syntactic information, it is unknown whether other lexical properties of a pronoun’s antecedent, such as its semantic and phonological features, are also retrieved during comprehension. Part of this uncertainty stems from the fact that evidence about the involvement of these features has been obtained using very different tasks and languages, and often in the domain of production, which renders problematic to draw inferences about the role of these features in comprehension. Below we summarize some of the previous studies,

point out several issues in their design and interpretation, and propose two experiments that aimed to produce a controlled comparison of the role of semantic and phonological information in the processing of pronouns in languages with and without syntactic gender.

Previous evidence of lexical retrieval

Evidence for semantic effects comes mostly from cross-modal priming studies done in English, in which participants listened to sentences containing pronouns and performed a visual lexical decision after hearing the pronoun (Leiman, 1982; Nicol, 1988; Shillcock, 1982). These studies found that responses were faster for words that were semantically related to the antecedent of the pronoun, relative to a semantically unrelated control word. For example, Shillcock (1982) auditorily presented sentences like (1) (the asterisks mark the points where a visual probe appeared on the screen in different trials).

- (1) The teacher* did not board the train, for the* simple reason that it/he* was not going to the South Coast of England.

SCHOOL/STREET

Participants performed a lexical decision to either the word “school” (semantically related to “teacher”) or the unrelated control word “street”, which was matched in frequency and length to “school”. At the offset of the pronoun “he”, a priming effect was obtained: lexical decisions were faster when participants responded to “school” than to “street”. When the pronoun was “it” instead of “he”, lexical decision times did not differ, showing that the priming effect for “school” in the “he” condition could not be attributed to residual priming from having

encountered the word “teacher” at the beginning of the sentence, and therefore, that reactivation of the antecedent is what led to priming.

Meanwhile, previous production studies performed in languages with syntactic gender have shown that pronouns also trigger access to phonological and frequency information about their antecedent. In a lexical decision during naming study in German (Schmitt, Meyer & Levelt, 1999) participants were asked to verbally describe two successive pictures of an object (e.g. a flower). When the two pictures showed the same object in different colors, participants typically used a pronoun to refer to the repeated object. Schmitt et al. found that if participants were interrupted to perform a lexical decision task when they were about to produce a pronoun, they showed reliable inhibition effects for words phonologically related to the antecedent. That is, words that shared the same onset with the antecedent of the pronoun (e.g., *Bluse* ‘blouse’ phonologically related to *Blume* ‘flower’) showed longer reaction times than unrelated control words matched in length and frequency.

Finocchiaro and Caramazza (2006) examined whether the lexical frequency of an antecedent noun would affect naming latencies in pronoun production in Italian. They hypothesized that if processing a pronoun involves re-activating the lexical entry of its antecedent, then naming latencies for utterances containing pronouns should be directly modulated by the frequency of the antecedent noun: pronouns with low-frequency antecedents should give rise to longer naming latencies than pronouns with high-frequency antecedents. In their naming task, participants were shown a written verb in the infinitive (e.g., “portare”, ‘to bring’) followed by a picture of an object (e.g., a helmet). They were asked to produce a simple imperative sentence

using the verb, a pronominal form corresponding to the name of the object and an indirect object, for example, “portamelo” (‘bring *it* to me’). Indeed, participants produced simple verb + clitic utterances more slowly when the noun replaced by the pronominal clitic was of low frequency than when it was of high frequency. Finocchiaro and Caramazza (2006) argued that their results supported the idea that pronouns engaged lexical access of their antecedent at a level where its grammatical features were represented.

Unresolved issues

So far, the evidence reviewed above supports a version of the lexical re-access hypothesis in which pronouns reactivate semantic, syntactic and phonological information about their antecedents in comprehension. However, there are several reasons that render this conclusion problematic. First, the finding of phonological inhibition effects in production by Schmitt and colleagues has not been replicated: for example, when Jescheniak, Schriefers & Hantsch (2001) used a picture-word interference paradigm, they failed to find any phonological effects on the generation of pronouns across four different experiments. This has raised the question of whether inhibition effects generalize to other production tasks.

Relatedly, all the evidence for phonological inhibition effects comes from studies of production, rather than comprehension. In production, since words have to be articulated, the connection between a word’s meaning and form features might be stronger, and retrieval of an antecedent’s syntactic representation might automatically activate its form as well. In comprehension, on the other hand, only the syntactic and

semantic properties of a pronoun's antecedent are necessary for interpretation. This could render retrieval of form properties such as its phonology and lexical frequency unnecessary in comprehension. In fact, at least one previous eye-tracking study in English failed to find any effects of antecedent lexical frequency (van Gompel and Majid, 2004). This could represent a cross-linguistic contrast between English and languages with syntactic gender, but it could also result from differences between the comprehension and production systems.

Furthermore, the interpretation of the semantic facilitation effect in cross-modal priming studies is problematic for several reasons. The first is that semantic association effects in these studies are highly dependent on the choice of control words. Inferences in these studies are drawn from the comparison between related and control words (e.g. "school vs. street" in the example above) at the point of interest (e.g. immediately after a pronoun), and control and related words are usually matched in either length and frequency or baseline reaction time. However, a study by McKoon, Ratcliff and Ward (1994) questioned the efficacy of this procedure in controlling for other lexical variables, or for the degree of fit between the control word and the sentence context. They provided an illustration by changing the type of control words, which they did not select separately but chose instead among the target words from the other experimental sentences. This ensured that lexical properties between control and related words were more closely matched, as the same words fulfilled both roles across the experiment. In four studies, McKoon & Ratcliff showed that this eliminated the semantic priming effect in test positions, and they further showed that with the usual matching procedure, this difference reappeared (for a

rejoinder, see Nicol, Fodor, & Swinney, 1994). These findings highlight that the choice of appropriate control comparisons is important to establish the existence of semantic priming effects.

A second problem is that semantic priming effects in cross-modal studies might be partly due to task-related strategies, as opposed to processes indexing automatic antecedent reactivation. This is because this paradigm encourages participants to develop task-related strategies, as detecting semantic relationships between words helps them perform well in the lexical decision task. It has been repeatedly shown that priming effects are sensitive to participants' strategies: for example these effects grow larger as the proportion of related words in a task increases (for review, see Neely, 1991). Therefore, it is unclear whether the previously obtained association effects actually reflect the automatic reactivation of a semantic antecedent representation in the lexicon, and also whether the results from the cross-modal priming paradigm can carry over to more naturalistic comprehension studies where participants are not assigned a specific task other than understanding the meaning of a sentence.

In sum, there are two main issues that emerge from previous findings. The first one is that it is unclear whether speakers of languages with syntactic gender access phonological and lexical semantic antecedent information in addition to syntactic information during comprehension. This leaves room for reasonable doubt over whether reaccessing the form of an antecedent should be seen as part of pronoun processing in general, or only as a necessary step in the production of anaphoric forms. Second, for languages without syntactic gender such as English, there is not

yet clear evidence that speakers access any type of lexical information, which has motivated proposals that only a discourse-based antecedent representation is reaccessed in English. In what follows, we outline two experiments that addressed these concerns by performing a side-by-side assessment of phonological and semantic effects using the same task and a closely controlled comparison across languages.

Overview of the experiments

We addressed the question of which type of antecedent representation is retrieved during coreference in reading comprehension. We focused on semantic and phonological properties about a pronoun's antecedent. We further asked whether the retrieval of these properties differs between languages that have syntactic gender, like German (Experiment 6), and those that don't, like English (Experiment 7). We used eye-tracking because it provides a more implicit measure of processing and obviates the need for participants to make conscious decisions about the relationships of interest.

We used a new strategy to probe for the nature of the antecedent representation reaccessed by pronouns. Our experimental design probed for the features of the antecedent by varying the type of relationship between the antecedent and the word immediately following the pronoun, which will be referred to as the *cohort* word:

(2) Semantic conditions

a. *Pronoun, Related*

The maintenance men told the **singer** about a problem. They had broken **his piano** and would have to repair that first.

b. *Pronoun, Unrelated*

- The maintenance men told the **deputy** about a problem. They had broken **his piano** and would have to repair that first.
- c. *Determiner, Related*
The maintenance men told the **singer** about a problem. They had broken **the piano** and would have to repair that first.
 - d. *Determiner, Unrelated*
The maintenance men told the **deputy** about a problem. They had broken **the piano** and would have to repair that first.

(3) Phonological conditions

- a. *Pronoun, Related*
The maintenance men told the **singer** there would be a delay. They said that **his sink** wouldn't be installed until next month.
- b. *Pronoun, Unrelated*
The maintenance men told the **deputy** there would be a delay. They said that **his sink** wouldn't be installed until next month.
- c. *Determiner, Related*
The maintenance men told the **singer** there would be a delay. They said that **the sink** wouldn't be installed until next month.
- d. *Determiner, Unrelated*
The maintenance men told the **deputy** there would be a delay. They said that **the sink** wouldn't be installed until next month.

In the semantic conditions, we adopted a manipulation that shared some features with the previous cross-modal priming studies. In the related conditions, the antecedent and cohort word shared a semantic/associative relationship (“singer-piano”), whereas in the unrelated conditions they did not (“deputy-piano”). In contrast with previous cross-modal priming studies, however, the target word, “piano”, was always held constant across the related and unrelated conditions, and relatedness was instead manipulated by varying the antecedent of the pronoun in the previous sentence. In previous eye-tracking studies, interlexical association between a semantic prime and a target word has been found to give rise to shorter reading times to the target word in both early and late reading measures, as well as an increased number of skips (Camblin, Gordon, & Swaab, 2007; Carroll & Slowiaczek, 1986; Morris & Folk, 1998).

In the phonological conditions, we adopted the same manipulation as the previously mentioned production study in German (Schmitt, Meyer & Levelt, 1999). In the related conditions, the antecedent and the cohort word shared the same onset (“singer-sink”), whereas in the unrelated conditions there was no phonological or orthographic overlap (“deputy-sink”). In previous eye-tracking studies, orthographic overlap has been found to give rise to inhibition effects (Patterson, Alcock, & Liversedge, 2011; Patterson, Liversedge, & Davis, 2009), producing longer reading times and higher skipping rates and regressions to a word when it is orthographically related to a preceding word. These inhibition effects have been attributed to slower recognition of the target word due to phonological competition from the prime and are consistent with slower decision times to orthographically related prime-target pairs in word recognition tasks (e.g., Bijeljac-Babic, Biardeau, & Grainger, 1997; Colombo, 1986; Davis & Lupker, 2006; Davis, Perea, & Acha, 2009; Davis & Taft, 2005; De Moor & Brysbaert, 2000; Grainger & Ferrand, 1994; Lupker & Coloumbo, 1994; Segui & Grainger, 1990).

Under a lexical re-access account, if comprehenders immediately reactivate semantic and/or phonological antecedent features upon encountering a coreferential pronoun, then we expect these features to impact eye fixations to the cohort word after the pronoun. If pronouns reactivate semantic information about their antecedents, then we predict a semantic priming effect in the pronoun conditions, with shorter reading times to the cohort word “piano” when the antecedent is “singer”, as compared with “deputy”. If pronouns reactivate phonological/orthographic antecedent features, then we predict a phonological

inhibition effect in the pronoun conditions, with longer reading times to the cohort word “sink” when the antecedent is “singer”, as compared with “deputy”. Importantly, if these effects are specifically due to referential processing (and not, for example, to residual activation from first encounter of the antecedent word), then they should be absent in the determiner conditions, as determiners should not re-activate the antecedent nouns.

4.3 Experiment 6: Antecedent representation in German

4.3.1 Methods

Participants

Participants ($n = 60$, mean age = 25 years, 46 females) were all native speakers of German and were recruited from the University of Potsdam community. All participants provided informed consent and received either course credit or payment for their participation.

Materials and design

The experimental materials consisted of 64 two-sentence item sets. The second sentence always contained a masculine or neuter possessive pronoun followed by the word of interest, the *cohort* word. The first sentence introduced the antecedent of the pronoun, which was realized as the direct object of a transitive verb. The antecedent was singular and of either masculine or neuter gender. In contrast, the sentential subject was always plural and feminine. This ensured that the possessive

pronoun in the second sentence unambiguously referred to the direct object of the previous sentence. A sample item set is shown in Table 1.

We varied whether the antecedent of the pronoun was related or unrelated to the word following the pronoun. This relationship could be semantic (e.g. *Zeichenlehrer-Bild*, ‘drawing teacher-painting’) or phonological/orthographic (e.g. *Zeichenlehrer-Zeitung*, ‘drawing teacher-newspaper’). In the latter case, the antecedent and the cohort word overlapped in at least the first two phonemes/characters of the word’s onset ($\text{mean}_{\text{phon}} = 2.51$, $\text{SD}_{\text{phon}} = 1.01$; $\text{mean}_{\text{orth}} = 2.88$, $\text{SD}_{\text{orth}} = 0.95$)⁴. For the unrelated conditions, the antecedent of the pronoun was replaced with a word that did not share a semantic or phonological relationship with the cohort word (e.g. *Administrator-Bild*, ‘administrator-painting’ and *Administrator-Zeitung*, ‘administrator-newspaper’ respectively). The related and unrelated antecedents were matched in lemma log frequency ($\text{mean}_{\text{rel}} = 0.88$, $\text{SD}_{\text{rel}} = 0.66$; $\text{mean}_{\text{unrel}} = 0.86$, $\text{SD}_{\text{unrel}} = 0.65$) and length ($\text{mean}_{\text{rel}} = 9.08$, $\text{SD}_{\text{rel}} = 2.92$; $\text{mean}_{\text{unrel}} = 9.30$, $\text{SD}_{\text{unrel}} = 2.85$) using the German WebCelex database (Baayen, Piepenbrock, & Van Rijn, 1993). To isolate relatedness effects specifically due to referential processing from relatedness effects due to priming that stemmed from the sentence context, we also manipulated whether a pronoun or a determiner preceded the target word. This resulted in a 2 (related/unrelated) \times 2 (semantic/phonological) \times 2 (pronoun/determiner) design.

The 64 item sets were divided into 8 lists, such that each list contained exactly one version of each item and 8 items in each condition. Thus, each participant saw each item and each condition, but never saw more than one version of the same item.

⁴ Phonetic transcriptions were obtained from the WebCelex database.

The experiment also contained 72 two-sentence filler items of comparable length and complexity. Filler items contained other kinds of referential expressions and anaphors, such as female pronouns.

German items	
<i>Pronoun related</i>	/Die Nachbarinnen /mochten den Zeichenlehrer , der im obersten Stockwerk wohnte. /Sie fanden, /dass sein Bild , an dem/ er in /seiner Freizeit gearbeitet hatte und das jetzt im Hausflur hing, sehr gut geworden war./ <i>The neighbors liked the drawing teacher, who lived on the top floor. They thought that his painting, on which he had worked in his spare time and now hung in the hall, had become very good.</i>
<i>Semantic conditions</i>	<i>Pronoun unrelated</i> Die Nachbarinnen mochten den Administrator der im obersten Stockwerk wohnte. Sie fanden, dass sein Bild , an dem er in seiner Freizeit gearbeitet hatte und das jetzt im Hausflur hing, sehr gut geworden war. ... <i>administrator... his painting...</i>
	<i>Determiner related</i> Die Nachbarinnen mochten den Zeichenlehrer , der im obersten Stockwerk wohnte. Sie fanden, dass das Bild , an dem er in seiner Freizeit gearbeitet hatte und das jetzt im Hausflur hing, sehr gut geworden war. ... <i>drawing teacher... the painting...</i>
	<i>Determiner unrelated</i> Die Nachbarinnen mochten den Administrator der im obersten Stockwerk wohnte. Sie fanden, dass das Bild , an dem er in seiner Freizeit gearbeitet hatte und das jetzt im Hausflur hing, sehr gut geworden war. ... <i>administrator... the painting...</i>
<i>Phonological conditions</i>	<i>Pronoun related</i> /Die Nachbarinnen /mochten den Zeichenlehrer , der im obersten Stockwerk wohnte. /Sie gingen sicher, /dass seine Zeitung nicht /aus seinem/ Briefkasten geklaut wurde./ <i>The neighbors liked the drawing teacher, who lived on the top floor. They made sure that his newspaper was not stolen out of his mailbox.</i>
	<i>Pronoun unrelated</i> Die Nachbarinnen mochten den Administrator der im obersten Stockwerk wohnte. Sie gingen sicher, dass seine Zeitung nicht aus seinem Briefkasten geklaut wurde. ... <i>administrator... his newspaper...</i>
	<i>Determiner related</i> Die Nachbarinnen mochten den Zeichenlehrer , der im obersten Stockwerk wohnte. Sie gingen sicher, dass die Zeitung nicht aus seinem Briefkasten geklaut wurde. ... <i>drawing teacher... the newspaper...</i>
	<i>Determiner unrelated</i> Die Nachbarinnen mochten den Administrator der im obersten Stockwerk wohnte. Sie gingen sicher, dass die Zeitung nicht aus seinem Briefkasten geklaut wurde. ... <i>administrator... the newspaper...</i>

Table 14 Sample set of experimental items and delimited regions in the German experiment.

Procedure

Participants were tested individually and eye movements were recorded using a desktop-mounted EyeLink 1000 eyetracker (SR Research, Mississauga, Ontario, Canada) interfaced with a Lenovo Thinkpad PC. The sampling rate was 1000 Hz. Stimuli were displayed on a 22-inch EIZO LCD monitor. Sentences were presented in

14 pt. Times New Roman font. Participants were seated 62 cm from the computer screen. At this distance 4.2 characters subtended around 1° of visual arc. Viewing was binocular, but only the right eye was recorded. Each sentence was displayed on a single line.

The experiment was implemented using the Experiment Builder software (SR Research, Mississauga, Ontario, Canada). A calibration procedure was performed at the beginning of each testing session, and re-calibration was carried out between trials as needed. Before the experiment began, each participant was instructed to read for comprehension in a normal manner. The participant triggered the onset of each sentence by fixating on a reference point on the left edge of the computer screen where the first word of the sentence was to appear. Each participant read six practice items before the experimental items were shown. All experimental and filler items were followed by a yes/no comprehension question to ensure that participants were attending to the stimuli. Comprehension questions never referred to the referential dependency between the pronoun and its antecedent. Average accuracy was 93.5%. The order of experimental and filler items was pseudo-randomized such that each experimental item was preceded by at least one filler item. The entire experimental session lasted approximately 45 minutes.

Analysis

The initial stages of data analysis were carried out using Data Viewer (SR Research, Mississauga, Ontario, Canada). Fixations were adjusted vertically only in cases where an entire sequence of fixations comprising at least half of the line fell

above or below a line of text (i.e. fixations were never adjusted either horizontally or individually). In addition, trials were omitted from the analyses if a long duration track loss occurred at any time during a trial (e.g., if there was no data for half a line of text or more). No trials were excluded due to this. Lastly, fixations shorter than 40 ms or longer than 1000 ms were excluded.

The region of interest consisted of the word following the pronoun, the *cohort region*. In addition, we report results from the *antecedent region*, which consisted of the antecedent noun, and the *pronoun region*, which consisted of either the determiner or the pronoun together with the preceding complementizer “dass” (‘that’). Including fixations to the left of a region of interest is a common procedure for analyzing short regions, because it is assumed that short words such as pronouns are processed during a fixation close to the left of the word when they are skipped (Ehrlich & Rayner, 1983; Garrod, Freudenthal, & Boyle, 1994; van Gompel & Majid, 2004). Table 1 shows the division into regions for a sample item.

We analyzed several eye-tracking measures. Phonological and semantic effects have been previously found in early measures, although an earlier study also observed semantic effects in late processing measures (Camblin, Gordon, & Swaab, 2007). Therefore, we focused our analysis on early measures, but we also report *total time* (the sum of all fixations in a region) to describe any processing differences that occurred after comprehenders’ initial processing of the region of interest. For early measures, we report *single fixation* (the duration of readers’ first fixation in a region when it is the only fixation in the region), *first fixation* (the duration of readers’ first fixation in a region, provided that they did not previously fixate on subsequent text)

and *first pass reading times* (the sum of all fixations on a critical region before readers' leave it for the first time, either to the left or to the right). Also, given that pronouns have been found to elicit a large number of regressions (Ehrlich & Rayner, 1983; van Gompel & Majid, 2004), and that phonologically related words have previously been found to be skipped more often, we report the *probability of regression* and *probability of skipping* for the pronoun and cohort regions. Skips of a region in a particular measure were treated as missing data points.

Statistical analyses were carried out with R, an open source programming language and environment for statistical computing (R Development Core Team, 2014), using the *lme4* package (Bates, Maechler, Bolker & Walker, 2014). Binomial measures and comprehension accuracy were analyzed using mixed effects logistic regression (Jaeger, 2008). Reading times were first logged and then analyzed with a linear mixed effects model. P-values were computed using the Satterthwaite's approximation for denominator degrees of freedom with the *lmerTest* package (Kuznetsova, Bruun Brockhoff, & Bojesen Christensen, 2014).

The model used included fixed effects of *determiner type*, *relatedness* and their interaction. Both main effects were coded using orthogonal contrasts. For the *determiner type* factor, the mean of the pronoun conditions was compared with the mean of the determiner conditions; thus, a positive estimate indicates that the pronoun conditions were read more slowly than the determiner conditions. Similarly, for the *relatedness* factor, the mean of the unrelated conditions was compared with the mean of the related conditions; a positive estimate indicates that the unrelated conditions were read more slowly than the related conditions.

With regard to the random effects structure of the model, we followed current guidelines in the psycholinguistics literature (Barr, Levy, Scheepers, & Tily, 2013) and initially constructed a maximal model that included random intercepts and slopes for all fixed effects and their interaction. Then, to determine whether the inclusion of random slopes was necessary we compared this maximal model with a model with only by-subject and by-item random intercepts. We performed log-likelihood ratio tests (Baayen, Davidson, & Bates, 2008) in the critical region –the cohort region– and found that the maximal model did not provide a significantly better fit to the data than the intercept-only model in neither of the measures of interest (*semantic conditions*: $X^2_{(18)} = 8.30, p = 0.96$; *phonological conditions*: $X^2_{(18)} = 11.72, p = 0.81$)⁵. Therefore, we adopted the intercept-only model, and applied it to the remaining regions of analysis for consistency. We present the model estimates in log milliseconds ($\hat{\beta}$), their standard error, and t- and p-values in the tables below.

4.3.2 Results

Table 2 shows means and standard errors in the three regions of analysis across the reading time measures of interest in the semantic and phonological conditions. Table 3 shows the results of the mixed effects model for the logged reading times, and pairwise comparisons and binomial measures are reported in the text.

⁵ The values provided are the averaged X^2 and p -values across all measures of interest.

		<i>Measures</i>			
		Single fixation	First fixation	First pass	
<i>Semantic conditions</i>					
Antecedent					
	Pron, Related	252 (7)	245 (5)	339 (11)	485 (16)
	Pron, Unrelated	250 (8)	246 (5)	366 (12)	515 (17)
	Det, Related	241 (6)	239 (5)	329 (11)	462 (17)
	Det, Unrelated	244 (7)	243 (4)	363 (12)	510 (19)
Pronoun					
	Pron, Related	286 (8)	250 (5)	382 (9)	478 (12)
	Pron, Unrelated	275 (7)	241 (4)	363 (8)	476 (12)
	Det, Related	263 (5)	242 (4)	320 (7)	391 (9)
	Det, Unrelated	256 (5)	246 (4)	343 (9)	425 (12)
Cohort					
	Pron, Related	222 (6)	226 (5)	311 (12)	353 (13)
	Pron, Unrelated	242 (6)	239 (5)	329 (11)	399 (14)
	Det, Related	228 (5)	234 (5)	288 (10)	340 (11)
	Det, Unrelated	223 (5)	234 (5)	288 (10)	365 (15)
<i>Phonological conditions</i>					
Antecedent					
	Pron, Related	249 (7)	245 (5)	337 (11)	465 (14)
	Pron, Unrelated	259 (7)	258 (5)	367 (12)	504 (18)
	Det, Related	236 (6)	237 (4)	322 (10)	464 (15)
	Det, Unrelated	241 (7)	246 (4)	358 (12)	515 (17)
Pronoun					
	Pron, Related	278 (8)	244 (4)	371 (9)	463 (12)
	Pron, Unrelated	271 (7)	247 (4)	382 (9)	469 (11)
	Det, Related	264 (6)	247 (4)	330 (8)	425 (12)
	Det, Unrelated	265 (6)	247 (4)	346 (9)	419 (10)
Cohort					
	Pron, Related	249 (7)	246 (6)	317 (11)	372 (12)
	Pron, Unrelated	236 (7)	242 (5)	319 (10)	365 (11)
	Det, Related	237 (7)	241 (5)	319 (10)	395 (13)
	Det, Unrelated	239 (6)	242 (5)	322 (11)	378 (15)

Table 15 Region averages and standard errors in milliseconds in the German experiment.

<i>Measures</i>																	
	Single fixation				First fixation				First pass				Total time				
	β	SE	t	p	β	SE	t	p	β	SE	t	p	β	SE	t	p	
<i>Semantic conditions</i>																	
Antecedent	Determiner type	0.01	0.02	0.55	0.58	0.01	0.02	0.55	0.58	0.01	0.02	0.37	0.72	0.03	0.03	1.30	0.19
	Relatedness	0.00	0.03	-0.15	0.88	0.01	0.02	0.62	0.54	0.08	0.02	0.00	<0.01	0.08	0.03	3.30	<0.01
	Det \times Rel	-0.03	0.05	-0.68	0.50	-0.01	0.03	-0.33	0.74	-0.01	0.05	-0.28	0.78	-0.02	0.05	-0.38	0.71
Pronoun	Determiner type	0.07	0.02	3.68	<0.01	-0.01	0.01	-0.43	0.67	0.11	0.02	5.59	<0.01	0.17	0.02	8.38	<0.01
	Relatedness	-0.02	0.02	-1.21	0.23	-0.01	0.01	-0.45	0.66	0.01	0.02	0.37	0.71	0.03	0.02	1.28	0.20
	Det \times Rel	-0.01	0.04	-0.26	0.80	-0.04	0.03	-1.49	0.14	-0.10	0.04	-2.54	<0.05	-0.07	0.04	-1.67	0.09
Cohort	Determiner type	0.01	0.02	0.57	0.57	-0.01	0.02	-0.63	0.53	0.07	0.02	2.87	<0.01	0.04	0.02	1.78	0.08
	Relatedness	0.04	0.02	2.02	<0.05	0.02	0.02	1.41	0.16	0.04	0.02	1.55	0.12	0.08	0.02	3.21	<0.01
	Det \times Rel	0.09	0.04	2.08	<0.05	0.05	0.04	1.52	0.13	0.07	0.05	1.55	0.12	0.10	0.05	1.98	<0.05
<i>Phonological conditions</i>																	
Antecedent	Determiner type	0.05	0.02	1.91	0.06	0.04	0.02	2.32	<0.05	0.03	0.02	1.47	1.47	0.00	0.03	0.02	0.98
	Relatedness	0.04	0.02	1.53	0.13	0.04	0.02	2.74	<0.01	0.08	0.02	3.58	<0.01	0.07	0.03	2.95	<0.01
	Det \times Rel	0.01	0.05	0.25	0.80	0.01	0.03	0.38	0.70	-0.01	0.05	-0.25	0.80	-0.04	0.05	-0.86	0.39
Pronoun	Determiner type	0.03	0.02	1.30	0.19	-0.01	0.02	-0.70	0.49	0.11	0.02	5.04	<0.01	0.11	0.02	5.61	<0.01
	Relatedness	-0.01	0.02	-0.43	0.67	0.00	0.02	-0.11	0.91	0.03	0.02	1.56	0.12	0.00	0.02	0.09	0.93
	Det \times Rel	-0.02	0.04	-0.54	0.59	0.01	0.03	0.20	0.84	-0.03	0.04	-0.61	0.54	0.00	0.04	-0.08	0.94
Cohort	Determiner type	-0.01	0.02	-0.47	0.64	-0.01	0.02	-0.44	0.66	-0.01	0.02	-0.48	0.63	-0.02	0.02	-0.95	0.34
	Relatedness	-0.02	0.02	-0.84	0.40	-0.01	0.02	-0.35	0.73	-0.01	0.02	-0.29	0.77	-0.04	0.02	-1.81	0.07
	Det \times Rel	-0.09	0.05	-1.82	0.07	-0.04	0.04	-1.05	0.30	0.00	0.05	0.09	0.93	0.03	0.05	0.57	0.57

Table 16 Statistics for reading time measures in the German experiment. Reliable effects are in bold font.

Semantic conditions

Antecedent region

No main effects or interactions were observed in this region with the exception of a main effect of relatedness in first pass and total time: unrelated antecedents were read more slowly than related antecedents. As comprehenders could not have computed relatedness until the cohort region, the relatedness effect might have been due to differences in the length or frequency of the antecedents, if our attempt to control from them was not fully successful.

Pronoun region

The pronoun region was skipped on 8.7% of trials. In both early and late measures, there was a main effect of determiner type: pronouns were read more slowly than determiners in single fixation, first pass and total time, and they were also skipped less often ($\hat{\beta} = -0.78$, $SE = 0.19$, $z = -4.18$, $p < 0.01$). These effects are not surprising the pronoun “his” is less frequent than the definite determiner “the”, and encountering a pronoun presumably engages additional cognitive processes, such as the search of an antecedent or the establishment of coreference. In addition, there was an interaction between determiner type and relatedness in first pass reading times; pairwise comparisons revealed that this was due to the determiners in the unrelated conditions being read more slowly than in the related conditions ($\hat{\beta} = 0.06$, $SE = 0.03$, $t = 2.14$, $p < 0.05$). There was no difference in the pronoun conditions ($\hat{\beta} = -0.05$, $SE = 0.03$, $t = -1.54$, $p = 0.12$).

Cohort region

The cohort region was skipped on 30.4% of trials. As in the pronoun region, there was a main effect of determiner type in first pass reading times and in the probability of regression measure: cohort words following pronouns were read more slowly and elicited fewer regressions than cohort words following determiners. Additionally, there was a main effect of relatedness: unrelated words elicited longer single fixation and total reading times, and also more regressions than related words.

Crucially, there was evidence that this relatedness effect was driven by the reading times in the pronoun conditions, as shown by a significant interaction between determiner type and relatedness in both single fixation and total time. Pairwise comparisons revealed that the relatedness effects were significant in the pronoun conditions but not in the determiner conditions: for pronouns only, when the cohort word was semantically related to the antecedent, it was read more slowly than when it was unrelated, in both single fixation (*pronoun*: $\hat{\beta} = 0.09$, $SE = 0.03$, $t = 2.74$, $p < 0.01$; *determiner*: $\hat{\beta} = -0.00$, $SE = 0.03$, $t = -0.07$, $p = 0.94$) and total time (*pronoun*: $\hat{\beta} = 0.13$, $SE = 0.04$, $t = 3.67$, $p < 0.01$; *determiner*: $\hat{\beta} = 0.02$, $SE = 0.03$, $t = 0.67$, $p = 0.50$).

The same pattern was observed in the remaining early measures. Since it was motivated by our theoretical hypothesis, we performed additional pairwise comparisons. As expected, semantically related cohort words were read more quickly than unrelated words in the pronoun conditions but not in the determiner conditions in first fixation (*pronoun*: $\hat{\beta} = 0.05$, $SE = 0.03$, $t = 2.07$, $p < 0.05$; *determiner*: $\hat{\beta} = -0.00$, $SE = 0.03$, $t = -0.08$, $p = 0.93$) and first pass reading times (*pronoun*: $\hat{\beta} = 0.07$, $SE = 0.03$, $t = 2.05$, $p < 0.05$; *determiner*: $\hat{\beta} = -0.01$, $SE = 0.03$, $t = -0.18$, $p = 0.86$). Figure 1 displays these effects as difference scores, which show the mean difference in reading times between the related and unrelated conditions ($\text{mean}_{\text{related}} - \text{mean}_{\text{unrelated}}$) for pronouns and determiners separately. Priming effects are shown by negative difference scores, which index that reading times were shorter for related than unrelated cohorts.

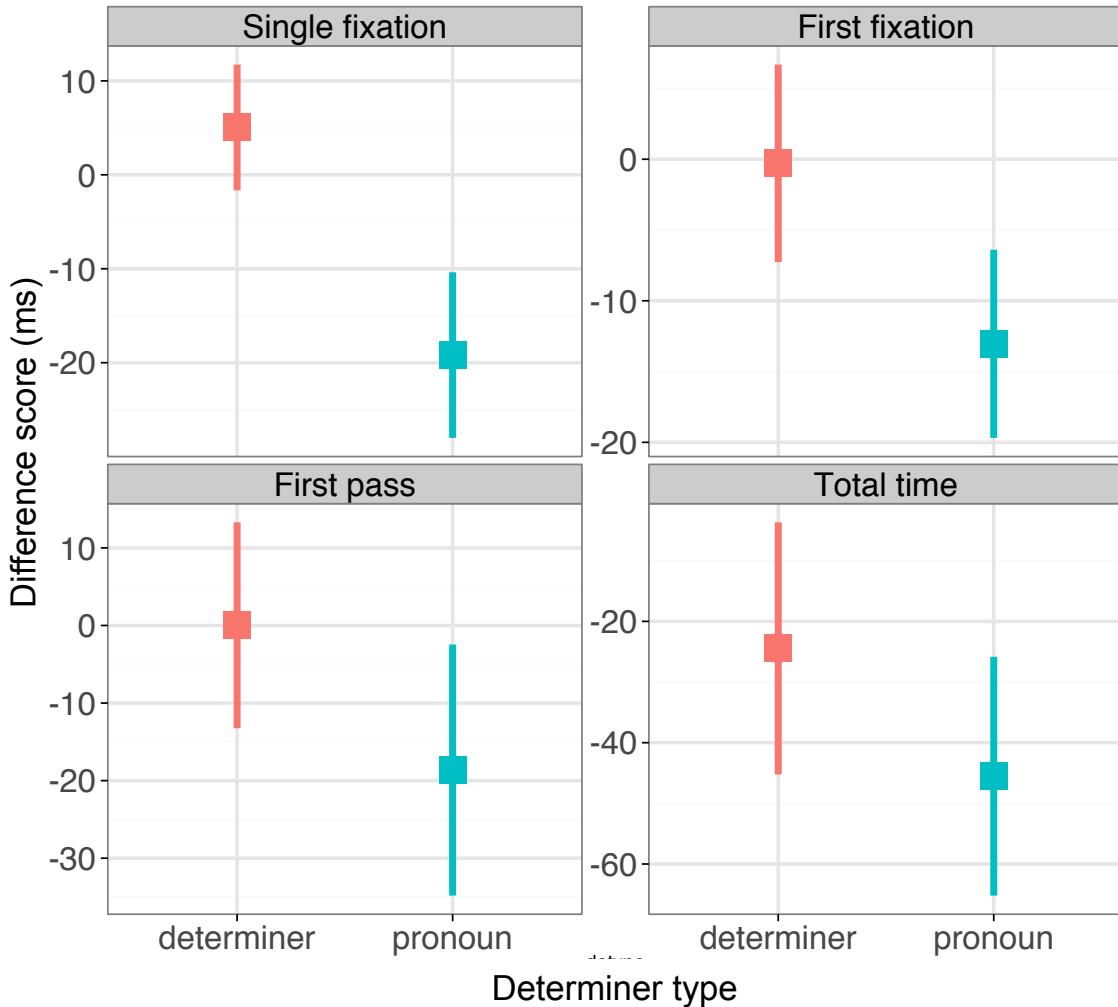


Figure 13 Semantic association effects in the German experiment. Mean difference scores and their standard error are shown with squares and bars respectively. Difference scores were computed as the mean difference between the related and the unrelated conditions. Therefore, negative difference scores index shorter reading times in the related than unrelated conditions (a priming effect).

Phonological conditions

Antecedent region

As in the semantic conditions, there was an unexpected main effect of relatedness in first fixation, first pass and total time: unrelated antecedents were read more slowly than related antecedents. This effect may have been due to residual differences in length or frequency between related and unrelated antecedents. Also, in

first fixation there was a main effect of determiner type, with the pronoun conditions being read more slowly than the determiner conditions.

Pronoun region

The pronoun region was skipped on 9.4% of trials. Only a main effect of determiner type was observed in this region: pronouns were read more slowly than determiners in first pass and total time, and they were also skipped less often ($\hat{\beta} = -0.39$, $SE = 0.17$, $z = -2.28$, $p < 0.05$).

Cohort region

The cohort region was skipped on 26.4% of trials. The only indication of a phonological interference effect specific to the pronoun conditions was a marginal interaction between determiner type and relatedness in single fixation duration. As this effect was motivated by our theoretical hypothesis we performed follow-up pairwise comparisons, but these failed to reveal any significant effect of relatedness in either the pronoun ($\hat{\beta} = -0.05$, $SE = 0.03$, $t = -1.46$, $p = 0.15$) or the determiner conditions ($\hat{\beta} = 0.02$, $SE = 0.03$, $t = 0.73$, $p = 0.47$). Therefore, the pattern of result suggests no effect of the antecedent's form on the reading times to the cohort word.

In addition, we ran a complimentary analysis in the region immediately after the cohort region. This was motivated by the fact that previous eye-tracking studies have reported inhibition effects in post-target regions (Patterson, Alcock, & Liversedge, 2011; Patterson, Liversedge, & Davis, 2009). The *cohort+1* region

showed a main effect of relatedness that was significant in single fixation ($\hat{\beta} = -0.06$, $SE = 0.02$, $t = -3.19$, $p < 0.01$), first fixation ($\hat{\beta} = -0.04$, $SE = 0.02$, $t = -2.78$, $p < 0.01$) and total time ($\hat{\beta} = -0.03$, $SE = 0.02$, $t = -2.87$, $p < 0.01$) and marginal in probability of regression ($\hat{\beta} = -0.26$, $SE = 0.15$, $z = -1.75$, $p < 0.09$). This effect was due to longer reading times in the phonologically related conditions. Importantly, there was no interaction between relatedness and pronoun type, suggesting that the relatedness effect affected the pronoun and determiner conditions alike.

4.3.3 Discussion

The results from this experiment can be summarized straightforwardly. There was evidence of semantic priming at the cohort word immediately after the pronoun when the antecedent and the cohort shared a semantic/associative relationship. The comparison between the determiner and pronoun conditions showed that this priming effect was specific to pronouns, as indexed by an interaction between relatedness and determiner type. In contrast, there was no evidence of phonological inhibition effects at the cohort region. Phonological inhibition effects were only found in the *cohort+1* region, and they affected pronouns and determiners alike. Overall, these results suggest that in German, coreference may involve reactivation of semantic but not phonological information about a pronoun's antecedent.

With regard to the phonological conditions, the word in the *cohort+1* region showed longer reading times in early measures and a marginal number of increased regressions in the related conditions. This inhibition effect replicates previous eye-tracking studies. For example, in one of these studies participants read sentences

where a target word was preceded by a prime word that could be orthographically related or unrelated, such as “More time was allowed to *extract/justify* the extra information that was needed.” (Patterson, Alcock, & Liversedge, 2011). Participants showed increased skips to the target word, “extra”, and more regressions and longer first pass times to the post-target region, “information”, when the prime and target words were orthographically related (“extract-extra”). In the current study, we also observed inhibition effects in the form of longer fixation times and more regressions at the post cohort region. These effects are surprising given that, in contrast with previous studies, there was a longer linear distance between the antecedent and the cohort word, which appeared in different sentences. Importantly, the effects occurred in both the pronoun and determiner conditions, which suggests that they were due to the residual activation from the phonologically related antecedent, and not to reactivation of its form specifically due to the processing of coreference.

In contrast, the semantic effects that we found showed specificity to the pronoun conditions, as evidenced by a significant interaction between the type of determiner word and the relatedness factors. Pairwise comparisons showed that cohort words that shared a semantic relationship with a pronoun’s antecedent were read faster than unrelated words across early and late measures. In contrast, there were no significant effects of semantic relatedness in the determiner conditions. These results support the hypothesis that German speakers re-activate semantic antecedent information immediately upon reading a coreferential pronoun.

However, the present results also raise an interesting question: at which level of representation are antecedents reactivated in comprehension, such that they can

induce priming effects? The Introduction raised a distinction between lexical representations, which refer to features of words stored in long-term memory, and discourse-level representations, which are constructed on the fly and contain information about the actions and entities described in the sentence. One possibility is that semantic priming effects could be due to reactivation of the semantic features of the antecedent noun in the lexicon. These features could in turn pre-activate words associatively related to the antecedent noun, since the semantic relationship between words such as “singer” and “piano” is assumed to be due to these words being stored closely together or sharing an increased number of semantic links (e.g. Collins & Loftus, 1975; Forster, 1976; Levelt, Roelofs, & Meyer, 1999; Morton, 1979). As a result, readers should recognize the cohort word more quickly in the pronoun conditions due to its preactivation, resulting in faster reading times.

Alternatively, the semantic effects that we found could be due to reactivation of a discourse-based antecedent representation. As comprehenders’ discourse model represents their interpretation of a sentence, it is expected to encode some kind of semantic information. It is less clear, however, what type of semantic information is encoded at the discourse level, and whether accessing that information should induce semantic priming effects. These effects could arise if comprehenders incorporate semantic associates of words in the sentence context to their discourse model. For example, when “singer” is encountered, comprehenders might add “piano” and “guitar” to their discourse representation, as well as other objects that are likely to be present in an event where a singer is present.

Therefore, the interpretation of the source of the semantic priming effects that we obtained will depend on the type of semantic information that comprehenders encode in their discourse model. Although this issue has not been directly addressed in the domain of noun relationships, some insights can be drawn from previous studies on implicit verb arguments. This research has proposed that creation of discourse roles for implicit arguments is not an automatic process, but is contingent, instead, on the strength of the lexical association between the verb and its arguments (Cotter, 1984; Corbett & Dosher, 1978; Garrod & Terras, 2000; McKoon & Ratcliff, 1981).

For example, Garrod & Terras (2000) addressed the question of whether, upon hearing the sentence “Keith wrote a letter yesterday”, speakers create a representation of the implicit instrument of the verb, ‘a pen’, in their discourse. They compared sentences where the relationship between the verb and its implicit argument was more or less canonical (e.g. “write-pen” vs. “write-chalk”) and found evidence of facilitated processing of the argument in subsequent reference. This facilitation, however, was only seen in the canonical case, supporting the claim that not all lexical relationships are automatically carried over to comprehenders’ discourse model. Instead, comprehenders might only create discourse roles for instruments that have strong lexical associations with the verb.

In sum, the present results support the idea that semantic information about a pronoun’s antecedent is reactivated immediately after encountering a coreferential pronoun. These findings, however, do no unambiguously address the type of antecedent representation that is reactivated in comprehenders’ memory. Specifically,

semantic priming effects could be explained by retrieval of either a lexical or a discourse-based antecedent representation. We tried to distinguish between these possibilities by conducting a study in English, a language where it has been previously suggested that coreference only involves discourse-based antecedent representations.

4.4 Experiment 7: Antecedent representation in English

In the present experiment we addressed whether semantic priming effects obtain in English reading comprehension. In English, in contrast with German, gender is not a syntactic property of nouns and morphosyntactic and lexical features such as animacy, number and gender have conceptual correlates, and thus it is reasonable to assume that they can also be represented at the discourse level. For example, the plural number of a noun usually correlates with the numerosity of its intended referent⁶, and gender in English is either semantic (e.g. “boy”, “king”) or stereotypical (“janitor”, “miner”). This has led to the proposal that referential pronouns in English only access a conceptual representation of their antecedent in a discourse model (the *discourse-only hypothesis*; Cloitre & Bever, 1988; Lucas, Tanenhaus & Carlson, 1990; Sag & Hankamer, 1984).

This claim is supported by both syntactic and processing evidence. From a syntactic perspective, Sag & Hankamer (1976, 1984) proposed a distinction between *surface anaphors*, such as ellipsis, which refer to a linguistic representation

⁶ Exceptions to this are nouns that are grammatically plural but conceptually singular (e.g. “trousers”, “scissors”), and nouns that are grammatically singular but are used to refer to multiple objects (e.g. “news”, “clothing”). These cases, however, are limited in English. In most other cases, there is a straightforward alignment between morphological and conceptual number.

associated with an antecedent, and *deep anaphors*, such as personal pronouns, which refer to some object in a model of the world constructed by the interpreter of the sentence or discourse. This distinction was based on several distributional differences, such as the fact that only surface anaphors require parallelism in syntactic form between antecedents and pronouns, or that whereas surface anaphors require a linguistic antecedent to be licensed, deep pronouns can be interpreted as referring directly to extrasentential elements.

There is also psycholinguistic evidence in support of a discourse-only account of antecedent representation in English. For example, earlier studies of coreference addressed whether the comprehension of pronouns was dependent on the linear distance between the pronoun and its antecedent. The reasoning was that if a linguistic or surface antecedent representation was necessary to establish coreference, then processing should become increasingly difficult as the distance between the antecedent and the anaphor increased. However, discourse factors such as focus and referential continuity were shown to affect the ease of interpreting these anaphors more than linear distance (Anderson, Garrod, & Sanford, 1983; Ehrlich & Johnson-Laird, 1982; Lesgold, Roth, & Curtis, 1979). In fact, distance effects are eliminated when focus is controlled for both definite noun phrase anaphors (Lesgold, Roth, & Curtis, 1979) and definite pronouns (Anderson, Garrod, & Sanford, 1983).

An additional source of evidence came from studies that aimed to isolate different levels of representation by comparing different tasks, with the hypothesis that availability of a particular level of representation should facilitate performance on tasks that required processing at that level (Cloitre & Bever, 1988; Lucas,

Tanenhaus & Carlson, 1990). Cloitre and Bever (1988) had participants make two different types of decisions (“word/non-word” vs. “abstract/concrete”) to an adjective modifying an antecedent noun immediately after the repeated antecedent or a coreferential pronoun were presented. Facilitation in the category and lexical decision task was compared to indicate the relative salience of either conceptual or surface information about the antecedent probe. They predicted that in the category decision, which was assumed to be more sensitive to processes at the discourse level, responses to probes following pronouns should show greater facilitation than following noun anaphors, if pronouns engaged processing at a discourse level. The reverse pattern should be true for participants’ lexical decisions, a task that was assumed to be initially more sensitive to lexical information. As predicted, they found that category judgements were faster after pronouns than noun anaphors, whereas lexical decisions were faster for noun anaphors than for pronouns (cfr. Lucas, Tanenhaus, & Carlson, 1990, for a similar distinction between naming and lexical decision).

Predictions

The evidence reviewed above supports the idea that coreference in English involves antecedent re-activation exclusively at the discourse level. This provides a useful case to examine what type of semantic information comprehenders encode in their discourse model. Specifically, our goal was to address whether the semantic priming effects found in German were due to the incorporation of lexical associates of the antecedent noun to comprehenders’ discourse model. If, as suggested by

previous evidence, English comprehenders only retrieve a discourse-based antecedent representation, then two possible scenarios follow. If semantic associates such as “piano” (when the antecedent is “singer”) are added to comprehenders’ discourse model, then similar semantic priming effects should be found across English and German. If, on the other hand, comprehenders do not add semantic associates to their discourse representation, then no priming effects should be found in English, in contrast with German.

An alternative possibility is for coreference in English to involve reactivation of both lexical and discourse-based antecedent representations. This would make the lexical re-access account a viable model for English. In this case, semantic priming effects should also be obtained in English, making it difficult to distinguish between the two scenarios outlined above. However, as lexical and discourse information might show a different time course in the processing of coreference (e.g. Garrod & Terras, 2000) the pattern of comprehenders’ eye-movements may be informative to distinguish between the two different sources of these effects. In addition, our design also included phonological conditions, with the aim of further probing for the existence of lexical re-access effects.

4.4.1 Methods

Participants

Participants ($n = 60$, mean age = 21 years, 38 females) were all native speakers of English and were recruited from the University of Maryland community.

All participants provided informed consent and received either course credit or payment for their participation.

Materials and design

We constructed 64 two-sentence item sets in a 2 (related/unrelated) \times 2 (semantic/phonological) \times 2 (pronoun/determiner) design. As with the German materials, the possessive pronoun was always in the second sentence and it was singular and had masculine gender. The first sentence introduced the antecedent of the pronoun, which was realized as the direct object of a transitive verb. In contrast, the sentential subject of the first sentence was always plural so it mismatched the pronoun in number. The antecedent of the pronoun varied in whether it shared a phonological or semantic relationship with the cohort word. Phonologically related antecedents overlapped with the cohort word in at least the first two phonemes/characters of the word's onset ($\text{mean}_{\text{phon}} = 3.33$, $\text{SD}_{\text{phon}} = 0.62$; $\text{mean}_{\text{orth}} = 2.64$, $\text{SD}_{\text{orth}} = 1.03$)⁷. Related and unrelated antecedents were controlled in log frequency ($\text{mean}_{\text{rel}} = 2.75$, $\text{SD}_{\text{rel}} = 0.55$; $\text{mean}_{\text{unrel}} = 2.72$, $\text{SD}_{\text{unrel}} = 0.61$) and length ($\text{mean}_{\text{rel}} = 7.41$, $\text{SD}_{\text{rel}} = 1.56$; $\text{mean}_{\text{unrel}} = 7.13$, $\text{SD}_{\text{unrel}} = 1.60$) using the SUBTLEX database (Brysbaert & New, 2009; available from the English Lexicon Project, Balota, Yap, Cortese, Hutchison, Kessler, Loftis, Neely, Nelson, Simpson, & Treiman, 2007). The regions of interest for one condition are shown in (4) and a sample item set is shown in (2-3):

⁷ Phonetic transcriptions were obtained from the American pronunciation entries of the Oxford Dictionary and revised by a native speaker of American English from the Maryland area.

- (4) /The maintenance men/ told the /singer/ about a problem. /They had brok/en **his**/ piano/ and would/ have/ to repair that first. /

The 64 item sets were divided into 8 lists, such that each list contained exactly one version of each item and 8 items in each condition. Thus, each participant saw each item and each condition, but never saw more than one version of the same item. The experiment also contained 72 two-sentence filler items of comparable length and complexity, which were adapted from the English fillers.

Procedure

Participants were tested individually, and eye movements were recorded using an EyeLink 1000 eyetracker (SR Research, Mississauga, Ontario, Canada), interfaced with a Dell PC. The sampling rate for recordings was 1000 Hz. Stimuli were displayed on a 23-inch Dell LCD monitor. Participants were seated 96.52 cm from the computer screen. At this distance 6.02 characters subtended around 1° of visual arc. The angular resolution of the eyetracker is 10–30 min of arc. Viewing was binocular, but only the right eye was recorded. Sentences were presented in 12 pt. fixed width Courier font. Each sentence was displayed on a single line.

The experiment was implemented using the Eye-Track software (<http://www.psych.umass.edu/eyelab/software/>). A calibration procedure was performed at the beginning of each testing session, and re-calibration was carried out between trials as needed. Before the experiment began, each participant was instructed to read for comprehension in a normal manner. The participant triggered the onset of each sentence by fixating on a reference point on the left edge of the computer screen where the first word of the sentence was to appear. Each participant read three

practice items before the experimental items were shown. Every experimental and filler item was followed by a yes/no comprehension question to ensure that participants were attending to the stimuli. Comprehension questions never referred to the referential dependency between the pronoun and its antecedent. Average accuracy was 94.2%. The order of experimental and filler items was randomized across participants. The entire experimental session lasted approximately 45 minutes.

Analysis

The initial stages of data analysis were carried out using Eye Doctor (<http://www.psych.umass.edu/eyelab/software/>). We applied the same processing criteria as in the previous experiment: fixations were only corrected vertically and only trials where there had been a long duration track loss were excluded. This resulted in the exclusion of 0.18% of all trials.

The same measures of interest and regions of analysis were used. As with the German materials, the pronoun region was lengthened by including the complementizer “that” before the pronoun. In items that did not have a complementizer ($n = 26$) the pronoun region was lengthened by including the last two characters of the preceding verb.

Statistical analyses were carried out with R using the same model as in the German experiment, which included fixed effects of *determiner type*, *relatedness* and their interaction and random by-subject and by-item intercepts. The intercept-only model was chosen after checking that the maximal model did not provide a significantly better fit to the data in the cohort region in neither of the measures of

interest (*semantic conditions*: $X^2_{(18)} = 8.37, p = 0.87$; *phonological conditions*: $X^2_{(18)} = 11.47, p = 0.85$). The model estimates in milliseconds ($\hat{\beta}$), their standard error, and t- and p-values are presented in the tables below.

4.4.2 Results

Table 4 shows means and standard errors in the three regions of analysis across the reading time measures of interest in the semantic and phonological conditions. Table 5 shows the results of the mixed effects model for the logged reading times, and pairwise comparisons and binomial measures are reported in the text.

		<i>Measures</i>			
		Single fixation	First fixation	First pass	Total time
<i>Semantic conditions</i>					
Antecedent					
	Pron, Related	253 (6)	250 (4)	312 (8)	489 (15)
	Pron, Unrelated	238 (7)	246 (4)	294 (7)	514 (16)
	Det, Related	273 (10)	253 (5)	305 (8)	517 (16)
	Det, Unrelated	254 (9)	249 (5)	295 (7)	529 (17)
Pronoun					
	Pron, Related	233 (6)	237 (5)	273 (7)	382 (14)
	Pron, Unrelated	246 (7)	236 (5)	280 (7)	416 (12)
	Det, Related	232 (7)	228 (6)	263 (8)	377 (12)
	Det, Unrelated	235 (7)	225 (4)	257 (7)	383 (11)
Cohort					
	Pron, Related	242 (6)	237 (4)	276 (8)	348 (13)
	Pron, Unrelated	247 (6)	241 (4)	279 (7)	383 (12)
	Det, Related	239 (5)	241 (4)	275 (7)	363 (11)
	Det, Unrelated	246 (7)	241 (5)	273 (7)	378 (11)
<i>Phonological conditions</i>					
Antecedent					
	Pron, Related	275 (11)	250 (5)	299 (7)	512 (16)
	Pron, Unrelated	254 (9)	245 (5)	303 (8)	512 (17)
	Det, Related	256 (7)	257 (5)	306 (8)	506 (17)
	Det, Unrelated	258 (7)	254 (5)	303 (8)	521 (18)
Pronoun					
	Pron, Related	239 (8)	237 (6)	287 (9)	404 (13)
	Pron, Unrelated	244 (8)	241 (5)	290 (9)	421 (14)
	Det, Related	223 (7)	223 (5)	257 (8)	399 (14)
	Det, Unrelated	215 (6)	225 (5)	253 (7)	388 (14)
Cohort					
	Pron, Related	251 (6)	244 (4)	286 (7)	418 (14)
	Pron, Unrelated	248 (7)	244 (5)	295 (8)	413 (14)
	Det, Related	244 (7)	246 (5)	287 (8)	437 (14)
	Det, Unrelated	250 (7)	249 (5)	301 (8)	434 (14)

Table 17 Region averages and standard errors in milliseconds in the English experiment.

<i>Measures</i>																
	Single fixation				First fixation				First pass				Total time			
	β	SE	t	p	β	SE	t	p	β	SE	t	p	β	SE	t	p
<i>Semantic conditions</i>																
Antecedent																
Determiner type	-0.03	0.03	-1.06	0.29	0.00	0.02	-0.13	0.90	0.01	0.02	0.41	0.68	-0.04	0.02	-1.54	0.12
Relatedness	-0.06	0.03	-2.17	<0.05	-0.01	0.02	-0.90	0.37	-0.03	0.02	-1.51	0.13	0.03	0.02	1.28	0.20
Det × Rel	0.01	0.06	0.22	0.83	0.00	0.03	0.05	0.96	-0.01	0.04	-0.38	0.71	0.03	0.05	0.68	0.49
Pronoun																
Determiner type	0.02	0.02	1.06	0.29	0.04	0.02	2.34	<0.05	0.06	0.02	2.84	<0.05	0.04	0.02	1.60	0.11
Relatedness	0.04	0.02	1.72	0.09	0.00	0.02	0.00	1.00	0.01	0.02	0.26	0.80	0.07	0.02	2.87	<0.05
Cohort																
Determiner type	0.01	0.02	0.37	0.71	-0.01	0.02	-0.66	0.51	0.00	0.02	-0.11	0.91	-0.02	0.02	-0.86	0.39
Relatedness	0.02	0.02	0.86	0.39	0.01	0.02	0.42	0.67	0.00	0.02	0.21	0.83	0.07	0.02	2.74	<0.01
Det × Rel	0.00	0.04	0.02	0.99	0.02	0.03	0.62	0.54	0.02	0.04	0.61	0.54	0.04	0.05	0.92	0.36
<i>Phonological conditions</i>																
Antecedent																
Determiner type	0.00	0.03	0.13	0.89	-0.04	0.02	-2.57	<0.05	-0.01	0.02	-0.60	0.55	0.00	0.03	0.20	0.85
Relatedness	-0.02	0.03	-0.83	0.41	-0.01	0.02	-0.79	0.43	0.00	0.02	-0.21	0.83	-0.01	0.03	-0.22	0.83
Det × Rel	-0.06	0.05	-1.06	0.29	-0.01	0.03	-0.25	0.80	0.01	0.04	0.16	0.88	-0.05	0.05	-1.04	0.30
Pronoun																
Determiner type	0.06	0.03	2.40	<0.05	0.04	0.02	2.43	<0.05	0.09	0.02	3.80	<0.01	0.05	0.03	1.76	0.08
Relatedness	0.00	0.03	-0.04	0.97	0.02	0.02	1.00	0.32	0.01	0.02	0.43	0.67	0.01	0.03	0.41	0.68
Cohort																
Determiner type	0.02	0.02	0.67	0.50	-0.01	0.02	-0.59	0.56	-0.01	0.02	-0.65	0.51	-0.04	0.03	-1.72	0.09
Relatedness	0.00	0.02	0.18	0.86	0.00	0.02	0.31	0.76	0.02	0.02	1.24	0.22	-0.01	0.03	-0.35	0.72
Det × Rel	-0.04	0.05	-0.81	0.42	-0.02	0.03	-0.56	0.58	-0.03	0.04	-0.65	0.52	-0.03	0.05	-0.60	0.55

Table 18 Statistics for reading time measures in the English experiment. Reliable effects are in bold font.

Semantic conditions

Antecedent region

No main effects or interactions were observed in this region with the exception of a main effect of relatedness in single fixation durations: unrelated antecedents were read more quickly than related antecedents. This effect may have been due to residual length or frequency differences between the antecedent words. However, it was not found across the other measures or in the phonological conditions (see below), so it may constitute a Type 1 error and we do not discuss it further.

Pronoun region

The pronoun region was skipped on 28.1% of trials. Early measures showed a main effect of determiner type: pronouns were read more slowly than determiners in first fixation and first pass times, and they were also skipped less often ($\hat{\beta} = -0.28$, $SE = 0.11$, $z = -2.48$, $p < 0.05$). Total reading times showed a main effect of relatedness: pronouns and determiners in the unrelated conditions showed longer reading times than in the related conditions. The only interaction that was found between pronoun type and relatedness occurred in the probability of regression measure. Pairwise comparisons showed that this was due to opposite effects of relatedness in the pronoun and determiner conditions: pronouns followed by semantically related cohort words were read more slowly ($\hat{\beta} = -0.43$, $SE = 0.20$, $z = -2.23$, $p < 0.05$) than pronouns followed by unrelated words, and the converse was true for determiners ($\hat{\beta} = 0.46$, $SE = 0.20$, $z = 2.29$, $p < 0.05$). As is argued in the discussion, the direction of the effect is the opposite of what would be predicted according to a lexical re-access account and the fact the it affects both pronouns and determiners suggests that it might not be related to processes related to coreference (i.e. processes that affect both the reading of pronouns and definite determiners).

Cohort region

The cohort region was skipped on 21.9% of trials. The only effect in this region was observed in total time, where unrelated cohorts were read more slowly than

related cohorts. Figure 2 displays the reading time profiles for purpose of comparison with the German results.

To ensure that the lack of a priming effect specific to the pronoun conditions was not due to a delayed effect of semantic relatedness, two complimentary analyses were performed. In the first analysis, eye-tracking measures more reflective of late processing were computed for the cohort region. However, no significant main effects of relatedness or interactions between determiner type and relatedness were found in right bound, re-read or regression-path times. Secondly, the region after the cohort, the *cohort+1* region, was also analyzed. Consistently with the results at the cohort region, this analysis revealed a main effect of relatedness, due to longer total reading times ($\hat{\beta} = 0.06$, $SE = 0.02$, $t = 2.55$, $p < 0.05$) and more regressions ($\hat{\beta} = 0.44$, $SE = 0.16$, $z = 2.77$, $p < 0.01$) in the unrelated conditions. However, no significant interactions between relatedness and determiner type were obtained. Therefore, the present results suggest that the semantic relationship between the cohort word and the antecedent facilitates reading times, but that this effect is similar for pronouns and determiners.

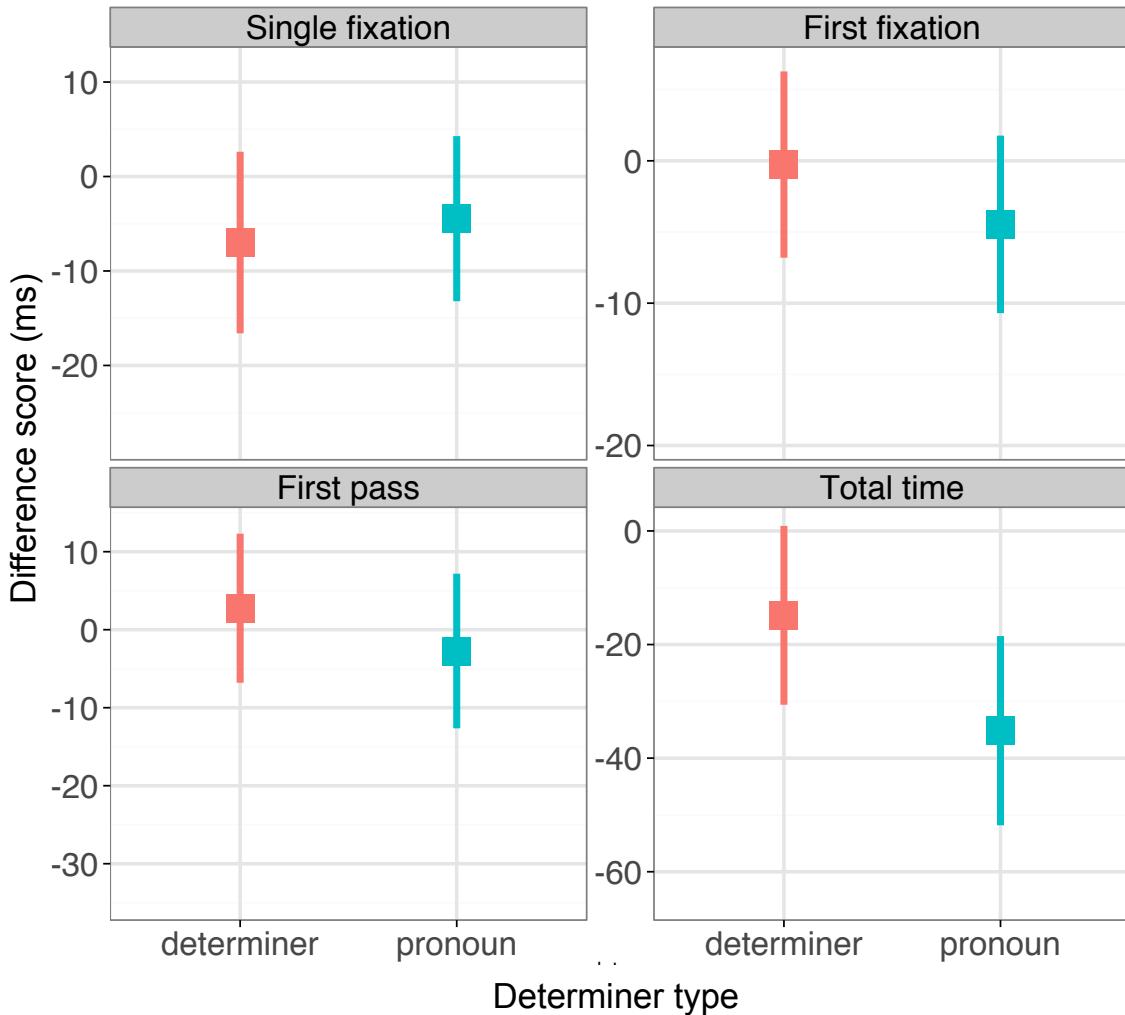


Figure 14 Reading time profiles in the English experiment. Mean difference scores and their standard error are shown with squares and bars respectively. Difference scores were computed as the mean difference between the related and the unrelated conditions. Therefore, negative difference scores index shorter reading times in the related than unrelated conditions (a priming effect).

Phonological conditions

Antecedent region

No effects or interactions were observed, with the exception of a main effect of determiner type, with the pronoun conditions being read more slowly than the determiner conditions.

Pronoun region

The pronoun region was skipped on 30.2% of trials. All early measures showed a main effect of determiner type: pronouns were read more slowly than determiners in single fixation, first fixation and first pass times, and they were also skipped less often ($\hat{\beta} = -0.26$, $SE = 0.11$, $z = -2.38$, $p < 0.05$).

Cohort region

The cohort region was skipped on 19.6% of trials. There were no main effects or interactions in the reading time measures. There was a marginal effect of determiner type was observed in the probability of regression measure: words that followed pronouns elicited more regressions than words that followed determiners ($\hat{\beta} = 0.23$, $SE = 0.12$, $z = 1.93$, $p = 0.05$).

As in the German study, we ran a supplementary analysis in the *cohort+1* region to probe for spillover effects. This analysis revealed a main effect of relatedness: in the related conditions, the spillover region was skipped less often ($\hat{\beta} = 0.26$, $SE = 0.12$, $z = 2.21$, $p < 0.05$) and elicited more regressions ($\hat{\beta} = -0.28$, $SE = 0.14$, $z = -2.09$, $p < 0.05$). Importantly, there was no interaction between determiner type and relatedness, suggesting that the inhibition effect affected the pronoun and determiner conditions alike.

4.4.3 Discussion

The goal of this study was to examine whether the semantic priming effects found in German were due to reactivation of lexical or discourse-based antecedent information. Our results show main effects of relatedness but not an interaction between these effects and the type of determiner word. This was the case in both the semantic and phonological conditions. The phonological conditions showed the expected inhibition effect, in that the post cohort region was skipped less often and elicited more regressions when the cohort word was phonologically related to the antecedent noun. This is the same profile that previous eye-tracking studies have found, and it shows that the phonological conditions were successful in eliciting form effects, and that the absence of effects specific to the pronoun conditions was not due to the materials that we used.

In contrast with the German results, although semantic relatedness clearly had an effect (as demonstrated by the main effect of relatedness in total times), pronouns did not show evidence of eliciting stronger antecedent semantic priming effects than determiners. This result held when additional eye-tracking measures were considered, and also when the post-cohort region was analyzed in order to probe for late or spillover semantic effects. A related concern is that the pronoun region in this experiment had a higher percentage of skips than in the German experiment (28.1% vs. 8.7%). If English comprehenders were less likely to read the pronoun, then antecedent reactivation may not have taken place in the pronoun conditions thus explaining the absence of semantic priming effects. However, we think that this explanation is unlikely for two reasons. The first one is that it relies on the

assumption that lack of fixations to a region implies lack of processing of that region; this does not follow, as short words are frequently processed parafoveally. Second, when we conducted a supplementary analysis that included only the trials where the pronoun region had been fixated, we obtained qualitatively similar patterns: only a main effect of relatedness and no interactions between relatedness and determiner type in either the cohort or the post-cohort regions were found.

In sum, the current results suggest that semantic associates of a pronoun's antecedent are not reactivated during coreference in English comprehension. This supports previous evidence that only a discourse-based antecedent representation may be reactivated in English, and builds on previous findings by showing that semantic associates of an antecedent noun do not seem to be included at the discourse level. We discuss the cross-linguistic implications of these findings below.

4.5 General Discussion

This study explored the nature of the antecedent representation that speakers reactivate upon encountering a coreferential pronoun. We asked whether this representation included semantic and phonological information, and we examined whether the retrieval of these features was language dependent by comparing German, a language with syntactic gender, and English, a language without it. We found that in reading comprehension German speakers show reliable semantic priming effects during coreference, whereas English speakers do not. In contrast with semantic priming effects, there was no evidence of phonological reactivation in neither of these languages. We examine each of these profiles in turn.

Phonological effects

The absence of phonological effects in both languages suggests that the form of a pronoun's antecedent is not reactivated in comprehension. In German, where we observed semantic priming effects, these results further suggest that lexical retrieval is selective, in that some kinds of information can be retrieved without others. This is consistent with models of the lexicon in comprehension where semantic, syntactic and phonological properties of words are represented separately (e.g., Forster, 1976; Morton, 1979) and with results from production, which have shown temporal dissociations (van Turennout, Hagoort, & Brown, 1997, 1998) and selective impairment in the retrieval of semantic and phonological information (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Garrett, 1980).

Importantly, the lack of phonological effects in our experiments is not likely due to lack of experimental power or a problem in the construction of the materials. Both experiments reliably showed phonological inhibition effects not specific to the pronoun conditions: in German, the post cohort region showed longer fixation times and more regressions; similarly, in English, the post cohort region was skipped less often and elicited more regressions in the related conditions. These inhibition effects replicate previous eye-tracking studies and show that the phonological manipulation in our materials was able to impact participants' eye movements. The fact that inhibition effects affected pronoun and determiners alike suggests, however, that they were not due to antecedent retrieval during coreference. Instead, a more likely explanation is that they resulted from residual activation of the antecedent noun.

When the antecedent word was read, its orthographic and phonological representations should have become activated. If these representations had not fully decayed by the time the cohort word was read, they could have interfered with recognition of the cohort word, resulting in the inhibition effects that we observed across the board.

The lack of phonological effects in comprehension contrasts with production findings by Schmitt, Meyer & Levelt (1999). These authors found inhibition effects in a lexical decision task during naming for words that shared an onset with the antecedent of a pronoun. However, one concern is that it is unclear how robust these effects are, as a follow-up study that used a picture-word interference paradigm failed to replicate these results across four different experiments (Jescheniak, Schriefers & Hantsch, 2001). Another possibility that there is a real contrast between the comprehension and production of anaphoric forms, such that the form of an antecedent is reactivated when a pronoun is produced, but not when it is heard or read. Future work will be necessary to address this issue.

Semantic effects

Our results suggest that speakers of languages with syntactic gender retrieve semantic antecedent information immediately upon encountering a pronoun. However, this finding was compatible with two different interpretations. One possibility was that this effect was due to lexical re-access. This should be the case if re-activation of semantic antecedent information preactivates semantically or associatively related words in the lexicon, increasing their baseline activation levels.

Thus, comprehenders should have an easier time recognizing the cohort word, resulting in shorter fixation measures of early processing such as single fixation, first fixation and first pass time. However, semantic facilitation may have also been due to retrieval of a discourse-based antecedent representation. As a discourse model encodes how comprehenders represent the situation described by a sentence, it is expected to include semantic information. If semantic associates of the antecedent noun are incorporated to the discourse, the reactivation of a discourse representation may have resulted in the semantic priming effects that we observed.

This second possibility was addressed in the English experiment. In contrast with German, nouns in English do not have syntactic gender. Instead, in English gender is either conceptual or stereotypical and the other features that pronouns require for agreement (person, animacy and number) have clear correlates at the discourse-level. This fact has led previous accounts to propose that all the relevant information to establish coreference in English may be represented at the discourse level, obviating the need for lexical re-access (Cloitre & Bever, 1988; Lucas, Tanenhaus & Carlson, 1990; Sag & Hankamer, 1984).

Consistently with these accounts, our results failed to find evidence of semantic priming effects in English comprehension. This suggests two things. The first one is that comprehenders may not automatically incorporate semantic associates of an antecedent noun to their discourse model. A similar conclusion has been reached by studies about implicit verb arguments (Cotter, 1984; Corbett & Dosher, 1978; Garrod & Terras, 2000; McKoon & Ratcliff, 1981), which have found that the creation of discourse roles for these arguments is contingent on the strength of the

lexical association between them and the verb. The second conclusion is that semantic priming effects may be uniquely tied to gender reactivation. These effects might only be present in languages with syntactic gender, motivating a case of cross-linguistic variation in the type of antecedent representations that speakers retrieve during comprehension.

This conclusion is striking given earlier findings of semantic priming effects in English in cross-modal priming studies (Leiman, 1982; Nicol, 1988; Shillcock, 1982). We pointed out two possible concerns with these findings. The first one is that these semantic priming effects were obtained by comparing lexical decision times between a word semantically related to a pronoun's antecedent and a control word matched in length and frequency. This design is problematic since matching in only length and frequency may not adequately control for other lexical variables, or for differences in the degree of fit between the context (not just the antecedent) and the related and unrelated words. There have been demonstrations that changing the type of control word can make semantic priming effects appear or disappear (for discussion, see Nicol, Fodor, & Swinney, 1994 and McKoon & Ratcliff, 1994). In the current study, we addressed this issue by holding the target word constant across conditions, and varying the antecedent of the pronoun instead.

The second concern is that semantic effects in cross-modal priming tasks may not exclusively index automatic reactivation processes. Instead, this paradigm encourages participants to develop task-related strategies, as detecting the semantic relationship between the target word and the antecedent helps subjects perform well in the task. The contribution of conscious strategies to semantic priming effects has

been repeatedly shown in the lexical decision task (for review, see Neely, 1991). To address this issue, we used a more implicit measure of processing, and made sure that the relationships of interest were never highlighted or relevant to answer the comprehension questions in our experiments.

Our results show that semantic priming effects do not obtain in English when a more implicit paradigm is used. As with the phonological conditions, the lack of an effect in our experiment is unlikely to be due to insufficient experimental power or badly constructed materials. This is because we found reliable priming effects at the cohort region in total times, as well as less regressions and shorter total reading times in the related conditions at the *cohort+1* region. The fact that these effects were observed for both pronouns and determiners suggests that they were due to priming from the sentence context, and not from antecedent reactivation specifically due to coreference.

Coreference across languages

Finally, the broader motivation for these studies was to examine the flexibility of the mechanisms that are used to implement coreference across languages. We observed that in English, most of the information necessary to compute coreference can be represented at the discourse level. In contrast, languages like German, Spanish and Italian have syntactic gender. In many instances a noun's syntactic gender does not have conceptual correlates, as with inanimate words like "chair" or abstract words like "freedom". In other cases, the syntactic gender of a noun may be at odds with its

conceptual gender, such as in the word ‘girl’ in German, which has neuter syntactic gender although it clearly denotes female entities.

The issue then is whether this difference between the grammatical systems of English and German affects speakers’ processing in comprehension. If processing is maximally uniform across languages, then the sequence of steps that speakers engage during coreference should be the same in English and German. That is, English speakers should engage in antecedent lexical re-access, regardless whether the retrieval of lexical information is necessary in this language. Alternatively, if speakers processing is guided by the properties of their grammar, they might only retrieve the representations that are necessary to license a pronominal dependency in their native language. In this case, we should expect lexical re-access to take place in German, but not in English.

Our findings show that comprehenders might only engage in lexical retrieval when syntactic gender information is necessary to license the relationship between a pronoun and its antecedent. This suggests that in languages with syntactic gender, a lexical antecedent representation is almost immediately retrieved during reading. In addition, our results support the claim that the processing of pronouns varies cross-linguistically. Although this conclusion has been previously suggested (e.g., Meyer & Bock, 1999, see Garnham, 2001, for review), the present study shows demonstrates this point by using closely matched materials and the same design across languages.

4.6 Conclusion

Two eyetracking experiments explored the sensitivity of readers' eye movements to semantic and syntactic information about a pronoun's antecedent. In German, we found evidence for antecedent semantic reactivation in the absence of phonological effects. In English, there was no evidence of either semantic or phonological effects. This suggests that lexical re-access is implemented in comprehension in languages with syntactic gender, but that in English, only discourse information may be accessed during coreference. Taken together with previous findings, these results suggest that the retrieval of antecedent representations during coreference depends on the type of information relevant to the grammar of each language.

5 Conclusions

5.1 Specific conclusions

This dissertation examined how cross-linguistic variation in the availability of number morphology and syntactic gender affects memory retrieval during sentence comprehension. I reported results from several reading time studies on two long-distance dependencies: subject verb agreement and pronominal coreference. The study of these dependencies was part of a broader effort to examine the flexibility of the cognitive architecture across languages by studying how variation in number and gender features affects the nature of the representations and processes that speakers deploy in comprehension. In this section I report specific experimental conclusions and remaining questions that resulted from these studies. In the next section I outline their broader implications for the cognitive architecture of language processing.

5.1.1 Agreement

The agreement studies examined how morphological variation across languages impacts comprehension mechanisms by comparing the processing of subject-verb agreement in English and Spanish, a language with richer agreement morphology. I considered several ways in which a richer morphological system could affect memory retrieval. The results showed both cross-linguistic similarities and differences. They are summarized below, together with questions for further research.

5.1.1.1 Speakers of Spanish and English implement memory retrieval identically

With respect to ungrammatical agreement, Spanish comprehenders always showed robust attraction effects in comprehension. While attraction effects in Spanish and in English were similar in size and timing, grammaticality effects – i.e. the amount of slowdown produced as a result of encountering a sheer agreement violation– were much larger in Spanish than in English. Therefore, these results support the claim that while morphological richness makes comprehenders initially more sensitive to agreement mismatches, facilitation due to attraction is surprisingly similar across languages. The observed cross-linguistic similarity in the response to ungrammatical agreement thus supports the hypothesis that retrieval is engaged uniformly cross-linguistically.

5.1.1.2 Morphological variation affects the strength of predictions across languages

With respect to grammatical agreement, Spanish but not English comprehenders showed evidence of processing cost when agreement morphology was carried by an auxiliary verb and a plural attractor noun was present. We suggested that this processing cost was due to prediction errors. In relative clause configurations, both the subject noun within the relative clause and the main clause subject noun (the attractor) might generate a prediction about the number of an upcoming verb. When the attractor noun is plural and the relative clause subject is singular, they should generate predictions for verbs of different number. If these two number predictions can interfere with each other, revision upon encountering the

relative clause verb may be sometimes unsuccessful. As a result, participants may process singular relative clause verbs as errors, resulting in increased reaction times.

The current results are compatible with the claim that due to their richer nominal plural morphology, Spanish comprehenders have more available input and a higher payoff to make stronger number predictions about upcoming verbs, in a way that English comprehenders do not. This might encourage Spanish speakers to routinely make stronger plural number predictions and to be more unsuccessful in revising them, resulting in processing cost when these predictions appear (temporarily) violated.

5.1.1.3 Why should differences in morphology affect prediction more strongly than retrieval?

The agreement results raise an interesting question: why should differences in morphology affect predictive more strongly than retrieval mechanisms across languages? Under models where the availability and reliability of cues are hypothesized to affect comprehension processes (cfr. the Competition Model; MacWhinney, 1987; MacWhinney, Bates, & Kliegl, 1984), this question can be restated as: why was retrieval not affected by a richer morphology in the processing of subject-verb agreement?

One possibility is that the invariability of attraction effects across languages is due to the use of retrieval as a repair strategy. The type of attraction effects that were observed in ungrammatical agreement sentences might speak to the way in which speakers implement repair processes when confronted with input that violates their expectations. Presumably, repair processes should be equally necessary in Spanish

and English, as speakers of all languages encounter situations where a given stream of speech or text does not conform to their predictions, as is the case, for example, in speech errors and disfluencies (Garrett, 1980). In the context of subject verb-agreement, this hypothesis states that when faced with an agreement violation, speakers of both Spanish and English might engage in retrieval using all available morphological cues equally in order to license the number-marked verb, regardless of the general reliability of these cues in their language.

Under this view, repair processes might be implemented differently from normal comprehension processes and might not reflect the properties of speakers' native language. This predicts that a useful way to predict when languages should differ is to consider the nature of the comprehension mechanisms of interest: in contrast with predictive operations, repair mechanisms might reflect language-general strategies to deal with noisy input, and as such, they might be used universally and uniformly across languages.

5.1.1.4 Testing prediction strength across languages

A second question raised by the present findings concerns the status of number predictions and their differential strength across languages. In contrast with the increasing amount of evidence that shows that speakers predict upcoming words (e.g. DeLong, Urbach & Kutas, 2005; Federmeier & Kutas, 1999; see Kutas, DeLong & Smith, 2011 for review), there is little work on the prediction of grammatical features such as number and gender. In particular, although there has been work on gender predictions (e.g., Van Berkum, Brown, Zwitserlood, Kooijman & Hagoort,

2005; Wicha, Bates, Moreno, & Kutas, 2003), it is undetermined, for instance, whether syntactic and conceptual gender predictions are implemented similarly in processing. In addition, we don't know how syntactic feature predictions are affected by morphosyntactic variation across languages.

There are several ways to extend the current findings to address these issues. For instance, in order to support the claim that number predictions are stronger in morphologically richer languages, it is important to establish a causal relationship between agreement morphology and prediction strength. This could be done, for instance, by showing a correlation between stronger predictions and slower decay rates. This would entail finding evidence that in morphologically richer languages like Spanish number predictions decay more slowly than in English.

There is some previous evidence that prediction strength can affect the rate of decay of information in memory. For instance, in a recent study, Wagers & Phillips (2014) used filled-gap and plausibility effects to provide evidence that the projection of the lexical-semantic features of a wh-filler phrase (e.g. "which stones") decayed more quickly than syntactic subcategorization information. They showed that comprehenders were still able to detect filled-gap effects after the introduction of bi-clausal intermediate material, but that in contrast, plausibility effects grew substantially weaker, consistent with the idea that semantic information about the filler decayed more rapidly.

If a similar logic is applied to the study of agreement, then the increased strength of number predictions should allow Spanish speakers to detect number mismatches even when faced with increasing intermediate material. In contrast, the

strength of grammaticality violations should be much weaker in English as the length of intermediate material increases, under the hypothesis that English speakers make weaker number predictions that are therefore more susceptible to decay. In contrast with the current experiments, this design would allow a within-language comparison of prediction strength, as the reaction time correlates of the detection of agreement violations could be measured in the presence of one, two or three-clause intermediate material within Spanish and English respectively.

In addition, an exciting direction for further research would be to examine whether within-language morphological differences can lead to differential processing effects. For instance, plurality can be realized differently within the same language, as some plural morphemes are more salient in surface form than others (e.g. “bowl-bowls” vs. “fish-fishes”). Can a change in the salience of morphological markers affect the strength of speakers’ plural expectations within the same language? The outcomes of the two lines of research presented here should provide a first step in supporting the idea that morphological variation across- and within-languages can affect predictive mechanisms in language comprehension.

5.1.2 Coreference

In the eye-tracking studies on coreference I explored the nature of the antecedent representation that speakers reactivate upon encountering a pronoun. I asked whether this representation included semantic and phonological information, and examined whether the retrieval of these properties was language dependent by comparing German, a language with syntactic gender, and English, a language

without it. I found that only German speakers showed evidence of semantic reactivation, and that neither German nor English speakers displayed phonological effects during coreference. Some of the implications of these findings are reviewed below.

5.1.2.1 In contrast with syntactic and semantic information, there is no evidence of retrieval of form representations during coreference in comprehension.

Across languages, there was never evidence that the form of an antecedent was reaccessed during coreference. In Experiments 4 and 5, we never observed effects of the antecedent's lexical frequency, a variable that has been previously associated with retrieval of the form of a word (Jescheniak & Levelt, 1994). In Experiments 6 and 7 in English and German, overlap in phonological/orthographic form between the antecedent and the cohort word after the pronoun did not have a measurable impact on participants' eye movements. Overall, our results suggest that the retrieval process targets representations in memory based on semantic and syntactic features, but not based on the form, either orthographic or phonological, of words in memory.

These results are informative with regard to the nature of the retrieval cues used within the content addressable memory architecture framework adopted in this dissertation. Retrieval processes within this architecture have been posited to be content-based, but the type of "content" that can be used as a cue is still an open issue (see Lewis & Vasishth, 2005, for review). The current findings suggest that neither phonological nor orthographic information is used at retrieval during comprehension. Consistent with this assumption, previous work has not found phonological effects at

retrieval in sentence comprehension. While phonological effects are obtained in tasks such as immediate serial recall and other short-term memory tasks (Jones, Hughes, & Macken, 2007; Jones, Macken, & Nicholls, 2004; see Caplan, Waters, & Howard, 2012, for review), the phonological similarity of words does not seem to affect retrieval in sentence processing (Kennison, 2004; Obata, Lewis, Epstein, Bartek, & Boland, 2010). Our results support these previous findings by showing that the lack of phonological effects on retrieval extends to the processing of coreference.

5.1.2.2 Some languages retrieve both the referent and the lexical representation of a pronoun's antecedent; for others, only retrieval of the referent takes place

While German speakers showed reliable antecedent semantic effects during coreference, no priming effects specifically tied to pronouns were observed in English. Therefore, we proposed that semantic effects in German were due to retrieval of a lexical antecedent representation, as reactivation of the antecedent noun in the lexicon should have primed semantically associated nouns, resulting in facilitated processing of the cohort word. In contrast, we claimed that, as proposed by previous accounts, only the referent of a pronoun is retrieved for coreference in English. We further argued that if semantic associates of the antecedent noun are not represented at the discourse level then the lack of semantic effects at the cohort word in English is expected.

This interpretation agrees with the conclusions of the verbatim recall studies reviewed in the Introduction, which showed that while morphosyntactic features are quickly lost across sentence boundaries, semantic content is more durable (Jarvella, 1971; Potter & Lombardi 1998; Sachs, 1967). These studies thus showed a temporal

dissociation between formal and conceptual information, supporting the claim that lexical information decays in memory more quickly than conceptual information. Our experiments contribute to this conclusion by showing discourse and lexical representations can also dissociate in terms of the words that they prime (or don't prime) in memory.

An interesting implication of these findings concerns the flexibility of the cognitive architecture across languages. In this sense, the coreference studies give rise to a different conclusion than the agreement studies, where the morphosyntactic cues of the verb seemed to be used uniformly across languages, despite of their functional importance. In contrast, for coreference it was found that in English, a language that lacks syntactic gender, comprehenders showed no evidence of retrieving a lexical representation and thus they did not display any antecedent semantic effects. In German, as was argued in previous chapters, retrieval of syntactic gender is required to check number agreement during coreference. This retrieval, which involves re-access of the syntactic properties of the antecedent representation in the lexicon, is what was posited yielded re-access to the antecedent semantic properties as well. More loosely, it could be said that the antecedent semantic properties come "for free" due the way the lexicon is structured, or more specifically, due the close connections between syntactic and semantic information in long-term memory.

This pattern thus suggests that English comprehenders only retrieve the information that is strictly necessary to license coreference. In contrast, a more uniform view of the cognitive architecture would posit that the same types of representations (discourse and lexical) should be retrieved across languages,

regardless whether they are necessary or not in processing. That is, under a more rigid architecture, the connections between the discourse and the lexicon should be established uniformly across languages. However, these results suggest that comprehenders only retrieve the information that is necessary according to their native grammar to license a referential dependency. Therefore, at least for the case of coreference, the cognitive architecture seems to be able to accommodate cross-linguistic differences in terms of what type of information that is required to license a grammatical dependency.

5.1.2.3 Capturing the dissociation between discourse- and lexical-level information within a content-addressable memory architecture

These results raise an interesting challenge for models that assume a single memory store. Most recent frameworks that adopt a content addressable architecture are in fact aligned with this assumption. This applies to psychological accounts of this architecture, which reject the distinction between working and long-term memory (e.g. McElree, 2000, 2006), and to a popular implementation of this architecture within the ACT-R framework (Lewis & Vasishth, 2005):

Traditionally, the capacity of WM was thought to be an important constraint on cognitive processing and to provide a basis on which to characterize differences among individuals and special populations. However, recent work has appealed more to differences in control and automatic processes than to differences in storage capacity (...). Measures of retrieval speed appear to dovetail with both of these recent directions in emphasizing that the successful execution of complex cognitive operations may depend more on our ability to shunt information between focal attention and memory than on the existence of a temporary store. (McElree, 2006: 194)

The declarative memory component in ACT-R serves functionally as both a long-term memory (encompassing semantic and episodic memory) and a short-term working memory, although there is not a structural distinction between the two. It is useful to think of each item or chunk in declarative memory as a feature bundle that can have relations to other chunks. (Lewis & Vasishth, 2005: 379)

There are several ways to capture our findings within a content-addressable memory architecture. One possibility is to give up the assumption of a single memory store, and propose that while lexical information is stored in long-term memory, speakers represent discourse-level information in their working memory. Thus, working memory would be used to encode the representation of not only sentences but also larger chunks of text. Although most research within this architecture has focused on intra-sentential relations, work addressing relations that can span across sentence boundaries is necessary if this architecture is to be used to account for the processing of linguistic phenomena such as coreference, which involves memory retrieval but which usually takes place across sentences in connected speech. In this case, retrieval might not be necessary during coreference in English (as the discourse antecedent representations will be in working memory already) but it would still be required in languages like German, where a lexical antecedent representation needs to be reaccessed as well.

Alternatively, one could reject the distinction between different levels of representation, and assume instead that retrieval always takes place over content-based memory chunks such as the ones assumed within the ACT-R architecture. One possibility to account for the existence of semantic priming effects in German, it could be stated that what differs across languages is the strength of the feed-forward

connections between these memory chunks and the lexical items originally activated by words in a sentence. In German, retrieval of the chunks should automatically activate neighboring long-term representations of the corresponding nouns, whereas in English these connections would be weaker or non-existent.

Alternatively, it could be proposed that the memory chunks that speakers build in memory are tagged, such that chunks can be labeled as being discourse, sentence, or lexical representations. The challenge would then be to provide an account of the nature of the retrieval cues that inform speakers that retrieval should take place over discourse, lexical chunks (or both), as these cues are not given in the surface form of words, as opposed to number and gender cues. This solution, while maintaining the existence of a single memory store, would have to assume that different type of representations are kept in this store, and it would capture cross-linguistic variability in terms of differential links between memory chunks.

5.2 Overall conclusions

Agreement and coreference can be seen as two test cases in processing that allow the study of a single question about cross-linguistic variation: *how do systematic changes in grammatical features across languages affect the representations and processes involved in sentence comprehension?* In the case of agreement, I examined retrieval in a case where the availability and functional importance of number morphology varied across languages. In other words, I explored a case where the *content* of the representations was the same across languages (i.e. number features specified as singular or plural by the grammar) but

where the *importance* of number information differed across languages: in Spanish, due to a more variable word order and its null subject status, number morphology is more available and reliable, both in acquisition (Slobin 1985; Phillips 2010; Pinker 1984) and in adult processing (Kail, 1989; MacWhinney, 2001).

I found, however, that this type of variation did not affect how speakers performed retrieval across languages: both Spanish and English speakers displayed quantitatively and qualitatively similar attraction rates, suggesting that number cues are weighted similarly across languages. These results do not rule out the possibility that retrieval cues can receive different weights within a content-addressable memory architecture. In fact, there is compelling evidence that semantic and syntactic cues carry different weights at retrieval (Kush, 2014; Parker, 2014; Van Dyke & McElree, 2011). Our results suggest, however, that cue weights depend on the nature of the information carried by a cue (e.g. whether it is structural, semantic, categorical, etc.) but not on its reliability or functional importance in a language.

In contrast, English and German were chosen to study the processing of coreference because they differed in how they represent gender information, which is semantic in English but syntactic in German. It was found that semantic priming effects during antecedent retrieval only obtain in German, suggesting that the reactivation of semantic associates of an antecedent noun is tied to the retrieval of its lexical representation but not to the retrieval of its referent in a discourse model.

Due to the several differences that exist between number agreement and coreferential dependencies, multiple hypotheses are possible to explain why similar retrieval profiles were observed for number agreement, whereas different

representations were recovered during antecedent retrieval for anaphors. Here I will only suggest two ideas that I think are fruitful for further research. The first one is to explore the possibility that memory representations are affected by the content of the information represented in speakers' native language, but not by the reliability or functional importance of this information. In other words, number information, which is similarly specified across Spanish and English as a morphosyntactic feature with two privative values (singular/plural), will be accessed in the same way at retrieval and will cause similar interference profiles across languages. On the other hand, gender information, which is semantic in English but syntactic in German, will not be reaccessed during coreference in the same way across languages, resulting in the processing dissociations that we observed.

The second possibility was suggested above and it involves the hypothesis that a crucial factor in cross-linguistic variation concerns the type of operations under study. In the case of agreement, similarities in retrieval were observed in cases where speakers were confronted with number violations and had to engage in retrieval as a repair strategy. In contrast, in the pronoun experiments only grammatical and referentially unambiguous relationships between pronouns and their antecedents were tested. It is therefore possible that the differences observed between these dependencies were due speakers engaging into different retrieval types due to the grammatical or ungrammatical contexts in which these dependencies were tested, and not due to inherent differences between dependency types.

Further research will be needed, especially to test the processing of grammatical agreement or ungrammatical coreference across languages. As they currently stand,

the present results suggest that there are exciting regularities to explore in the domain of cross-linguistic variation in processing, in particular at the level of the memory representations and retrieval processes that speakers deploy across languages.

Appendices

Appendix A. Item sets for Experiments 1–3

Spanish materials (Experiments 1 & 2)

The experiments contained 48 item sets of 4 conditions each. In Experiment 1, main verbs in the past tense were used inside the relative clause (conditions a-d). In Experiment 2, verbal phrases consisting of a future auxiliary preceded by the preposition “a” and nonfinite verbs were used inside the relative clause (conditions e-h). Conditions were as follows:

- Condition a: main verb, singular attractor, grammatical
- Condition b: main verb, plural attractor, grammatical
- Condition c: main verb, singular attractor, ungrammatical
- Condition d: main verb, plural attractor, ungrammatical
- Condition e: auxiliary, singular attractor, grammatical
- Condition f: auxiliary, plural attractor, grammatical
- Condition g: auxiliary, singular attractor, ungrammatical
- Condition h: auxiliary, plural attractor, ungrammatical

An example of a full set:

- 1a. La nota que la chica escribió en la clase alegró a su amiga.
 - 1b. Las notas que la chica escribió en la clase alegraron a su amiga.
 - 1c. *La nota que la chica escribieron en la clase alegró a su amiga.
 - 1d. *Las notas que la chica escribieron en la clase alegraron a su amiga.
 - 1e. La nota que la chica va a escribir en la clase alegrará a su amiga.
 - 1f. Las notas que la chica va a escribir en la clase alegrarán a su amiga.
 - 1g. *La nota que la chica van a escribir en la clase alegrará a su amiga.
 - 1h. *Las notas que la chica van a escribir en la clase alegrarán a su amiga.
2. El/Los tesoro(s) que el pirata captur(ó/aron)/va(n) a capturar en la bahía escondí(a/an) una infinidad de joyas preciosas.
 3. La(s) joya(s) que el rey guard(ó/aron)/va(n) a guardar en el palacio estaba(n)/estará(n) bajo vigilancia a toda hora.
 4. La(s) canción(es) que la mujer cant(ó/aron)/va(n) a cantar para la audiencia era(n)/es(son) un éxito de los setentas.
 5. El/los postre(s) que el cocinero prepar(ó/on)/va(n) a preparar en el restaurante era(n)/es(son) ejemplo de lo mejor de la región.

6. La(s) silla(s) que el carpintero construy(ó/eron)/va(n) a construir para el comedor colaps(ó/aron)/colapsar(á/arán) bajo el peso del perro de la casa.
7. El/los espejo(s) que el ladrón rompi(ó/eron)/va(n) a romper en el museo perteneci(ó/eron) al rey de Inglaterra hace mucho años.
8. El/los cuarto(s) que la criada limpi(ó/aron)/va(n) a limpiar con mucha prisa olí(a/an)/huel(e/en) a cerveza y a ropa sucia.
9. La(s) huella(s) que el detective encontr(ó/aron)/va(n) a encontrar en la oficina incriminab(a/an)/incriminar(á/arán) al gobernador en el crimen.
10. El/los secreto(s) que el espía escuch(ó/aron)/va(n) a escuchar por el teléfono pas(ó/aron)/pasar(á/án) enseguida al cuartel general.
11. La(s) fiesta(s) que el profesor organiz(ó/aron)/va(n) a organizar en el salón tuv(o/ieron)/tendr(á/án) mucha popularidad entre los alumnos.
12. El/los contrato(s) que la abogada firm(ó/aron)/va(n) a firmar con una lapicera autorizab(a/an)/autorizar(á/án) la venta de la compañía.
13. La(s) ciudad(es) que el turista visit(ó/aron)/va(n) a visitar durante el verano tení(a/an)/tien(e/en) muchas recomendaciones en las guías de turismo.
14. El/los candidato(s) que el votante apoy(ó/aron)/va(n) a apoyar en la última elección incumpli(ó/eron)/incumplir(á/án) las promesas hechas durante la campaña electoral.
15. El/los libro(s) que la historiador estudi(ó/aron)/va(n) a estudiar en la biblioteca contení(a/an)/contien(e/en) varios capítulos sobre la guerra del Paraguay.
16. El/los partido(s) que el jugador gan(ó/aron)/va(n) a ganar en el último torneo atraj(o/eron)/atraer(á/án) la atención de todos los medios.
17. El/los tumor(es) que el médico examin(ó/aron)/va(n) a examinar en el hospital creci(ó/eron)/crecer(á/án) enormemente durante las semanas siguientes.
18. El/los idioma(s) que el alumno aprend(ó/ieron)/va(n) a aprender en la universidad requiri(ó/eron)/requier(e/en) mucho esfuerzo y práctica intensiva.
19. La(s) historia(s) que el maestro cont(ó/aron)/va(n) a contar en la escuela hablab(a/an)/habl(a/an) de los antiguos mayas.
20. La(s) tela(s) que el costurero cort(ó/aron)/va(n) a cortar con mucho cuidado vení(a/an)/vien(e/en) desde la India.

21. El/los documento(s) que el político escondi(ó/eron)/va(n) a esconder en el cajón tení(a/an)/tien(e/en) detalles sobre un arma muy poderosa.
22. El/los payaso(s) que el cumpleañero esper(ó/aron)/va(n) a esperar con mucha anticipación lleg(ó/aron)/llegar(á/án) media hora tarde.
23. El/los descubrimiento(s) que el científico explic(ó/aron)/va(n) a explicar en un artículo revolucionar(á/án) el entorno académico.
24. La(s) estrella(s) que el astrónomo observ(ó/aron)/va(n) a observar en el oeste resplandecí(a/an)/resplandec(e/en) de una manera impresionante.
25. El/los número(s) que el cliente marc(ó/aron)/va(n) a marcar en la mañana pertenec(e/en) a una empresa de aguas corrientes.
26. La(s) carta(s) que el aplicante mand(ó/aron)/va(n) a mandar para un nuevo trabajo lleg(ó/aron)/llegar(á/án) a tiempo a su destino en Francia.
27. La(s) máquina(s) que el ingeniero repar(ó/aron)/va(n) a reparar en el taller serví(a/an)/sirv(e/en) para cosechar diferentes tipos de trigo.
28. La(s) foto(s) que el fotógrafo tom(ó/aron)/va(n) a tomar de la catedral particip(ó/aron)/participar(á/án) en un concurso con un prestigioso premio.
29. La(s) pintura(s) que el estudiante reconoci(ó/eron)/va(n) a reconocer en la revista era(n)/es(son) obra de un gran artista español.
30. La(s) crítica(s) que el diputado realiz(ó/aron)/va(n) a realizar contra la Iglesia apareci(ó/eron)/aparecer(á/án) en todos los diarios.
31. El/los producto(s) que el comerciante vendi(ó/eron)/va(n) a vender en el extranjero tení(a/an)/tien(e/en) un defecto de fábrica.
32. El/los edificio(s) que el arquitecto diseñ(ó/aron)/va(n) a diseñar para el gobierno qued(a/an)/quedar(á/án) en las afueras de la ciudad.
33. La(s) sustancia(s) que el químico añadi(ó/eron)/va(n) a añadir a la botella caus(ó/aron)/causar(á/án) una gran explosión en el laboratorio.
34. La(s) sonata(s) que el pianista toc(ó/aron)/va(n) a tocar el martes pasado/próximo recibí(ó/eron)/recibir(á/án) una mención especial del jurado.
35. La(s) caja(s) que el amigo ofreci(ó/eron)/va(n) a ofrecer con buena disposición era(n)/(es/son) para la mudanza al nuevo departamento.

36. El/los aumento(s) que el jefe prometi(ó/eron)/va(n) a prometer a los empleados llegar(ó/aron)/llegar(á/án) el día de Navidad.
37. El/los juguete(s) que el abuelo compr(ó/aron)/va(n) a comprar en la tienda era(n)/es(son) un regalo para su nieto.
38. El/los poema(s) que el niño recit(ó/aron)/va(n) a recitar en la clase commovi(ó/eron)/commover(á/án) al resto de sus compañeros.
39. El/los evento(s) que el testigo describi(ó/eron)/va(n) a describir en la corte coincidí(a/an)/coincid(e/en) con los registros policiales.
40. El/los avión(es) que el mecánico prob(ó/aron)/va(n) a probar para el gobierno tení(a/an)/tien(e/en) un nuevo sistema de radares de gran seguridad.
41. La(s) rata(s) que el gato caz(ó/aron)/va(n) a cazar en el jardín viví(a/an)/viv(e/en) en el granero de la granja.
42. El/los reporte(s) que el periodista redact(ó/aron)/va(n) a redactar para el diario contení(a/an)/contien(e/en) un gran número de errores y calumnias.
43. La(s) regulación(es) que el manifestante repudi(ó/aron)/va(n) a repudiar delante de la alcaldía permití(a/an)/permit(e/en) el uso de derivados del tabaco.
44. La(s) montaña(s) que el alpinista escal(ó/aron)/va(n) a escalar en el invierno quedab(a/an)/qued(a/an) al oeste de la frontera con Chile.
45. El/los automóvil(es) que el conductor manej(ó/aron)/va(n) a manejar por la autopista tení(a/an)/tien(e/en) fallas en el sistema de frenos.
46. La(s) blusa(s) que la modelo escogi(ó/eron)/va(n) a escoger para el desfile tení(a/an)/tien(e/en) lentejuelas de todos los colores.
47. La(s) mermelada(s) que el viajero comi(ó/eron)/va(n) a comer con mucho gusto contení(a/an)/contien(e/en) higos, peras y membrillo.
48. La(s) cosecha(s) que el agricultor plant(ó/aron)/va(n) a plantar en el campo sucumbi(ó/eron)/sucumbir(á/án) ante una plaga de insectos.

English materials (Experiment 3)

All experimental items were adapted from the second and third experiments reported in Wagers, Lau & Phillips (2009). The experiment contained 48 sets of 4 conditions. The verbal phrase inside the relative clause was always in the past tense and contained an auxiliary ("was"/"were") followed by an adverb (e.g. "always") and a main verb in the participle form (e.g. "seeing"). Conditions were as follows:

- Condition a: singular attractor, grammatical
- Condition b: plural attractor, grammatical
- Condition c: singular attractor, ungrammatical
- Condition d: plural attractor, ungrammatical

An example of a full set:

- 1a. The runner that the driver was always seeing during the commute waved to say hi.
- 1b. The runners that the driver was always seeing during the commute waved to say hi.
- 1c. * The runner that the driver were always seeing during the commute waved to say hi.
- 1d. * The runners that the driver were always seeing during the commute waved to say hi.

2. The industrialist(s) that the activist was/were always criticizing during meetings put(s) profit above concern for the environment.
3. The player(s) that the coach was/were always praising very enthusiastically decided to leave the team.
4. The volunteer(s) that the director was always relying on at the shelter was/were very dedicated to her/their work.
5. The author(s) that the librarian was/were enthusiastically recommending as a good read(s) is/are very popular in India.
6. The pilot(s) that the smuggler was/were secretly employing was/were discreet and good at thinking fast.
7. The colleague(s) that the researcher was/were always working with at the university came up with a crazy but inventive idea.
8. The patient(s) that the doctor was/were regularly seeing at the hospital was/were undergoing an experimental treatment.
9. The marine(s) that the officer was/were angrily in front of the platoon always behaved clumsily during training.
10. The sculptor(s) that the donor was/were generously supporting on a monthly basis has held several acclaimed exhibitions.
11. The caterer(s) that the hostess was/were decidedly recommending to her friends is/are excellent but outrageously expensive.
12. The senator(s) that the voter was/were patiently listening to at the political event tend(s) to speak for too long.
13. The teenager(s) that the farmer was/were regularly hiring to pick fruit every summer work(s) pretty hard.
14. The midfielder(s) that the fan was/were frantically cheering for during the game is/are reliably calm under pressure.
15. The quarterback(s) that the recruiter was/were attentively watching from the bleachers might win (a) lucrative contract(s).
16. The publicist(s) that the politician was/were regularly using during the crisis has/have years of experience.
17. The tenant(s) that the landlord was/were always criticizing due to his/their laziness stay(s) out late and never mows/mow the lawn.

18. The skateboarder(s) that the kid was/were passionately cheering for always sign(s) autographs for fans.
19. The chef(s) that the gourmet was/were always visiting at the restaurant came up with a daring new recipe.
20. The firefighter(s) that the ranger was/were frequently advising at the national park has/have special training in forest fires.
21. The overseer(s) that the worker was/were always complaining about during meetings has/have been firing anyone who mentions unions.
22. The heiress(es) that the bachelor was/were relentlessly pursuing at the charity is/are clever and won't be fooled easily.
23. The actor(s) that the producer was/were constantly calling to the set is/are late most mornings during shooting.
24. The accountant(s) that the administrator was/were always depending on to balance the books is/are very well paid.
25. The orphan(s) that the nun was/were regularly tutoring in algebra lost his/their parents in the ongoing civil war.
26. The translator(s) that the diplomat was/were desperately needing for communication is/are trying to facilitate peace talks.
27. The dancer(s) that the host was/were enthusiastically applauding at the gala has/have been performing since she/they was/were very young.
28. The supervisor(s) that the counselor was/were directly reporting to on a monthly basis provided advice and support.
29. The commentator(s) that the viewer was/were attentively watching every night has/have (a) very slick hairstyle(s).
30. The journalist(s) that the editor was/were always praising during meetings turn(s) her/their stories in promptly.
31. The customer(s) that the waitress was/were angrily looking at during her shift is/are the one(s) who leave(s) a lousy tip.
32. The therapist(s) that the survivor was sometimes meeting with after deployment is/are trying to help her/them recover from her/their awful ordeal.
33. The cheerleader(s) that the choreographer was/were frequently working with before the game train(s) all year for the national competition.
34. The cowboy(s) that the rancher was/were heatedly arguing with all the time has/have thirty years of experience on the range.
35. The surgeon(s) that the nurse was/were always assisting during surgery has/have an impeccable record in the operating room.
36. The bartender(s) that the patron was/were secretly gossiping about after work has/have (a) wild private life/lives.
37. The columnist(s) that the reader was/were angrily complaining about on a regular basis give(s) extreme opinions without backing them up.
38. The anthropologist(s) that the villager was/were always mentioning to his friends take(s) great pains to learn the local customs.
39. The bricklayer(s) that the contractor was/were always hiring for important projects doesn't/don't show up on time.
40. The comedian(s) that the prisoner was/were frequently watching on the TV always put(s) on a great show.

41. The criminal(s) that the judge was/were fairly sentencing at court is/are well known for hooliganism in the town.
42. The student(s) that the bully was/were cruelly teasing during lunch is/are too afraid to tell the teachers.
43. The philosopher(s) that the scientist was loudly condemning during the radio program rarely express(es) his/their ideas in plain language.
44. The lawyer(s) that the candidate was/were regularly consulting during the campaign is/are known for giving excellent legal counsel.
45. The musician(s) that the reviewer was/were highly praising in every article will probably win a Grammy.
46. The receptionist(s) that the boss was/were always eyeing never fail(s) to do a stellar job.
47. The actress(es) that the designer was/were kindly complimenting on every occasion is/are extremely fickle and demanding.
48. The programmer(s) that the manager was/were always overseeing at the software firm is/are eccentric and difficult to control.

Appendix B. Omnibus repeated measures analysis of variance

Results from a complementary analysis computing ANOVAs by-subjects (F_1) and by-items (F_2) using the ez package in R (Lawrence, 2013). MinF' statistics (Clark, 1973; Raaijmakers, Schrijnemakers, & Gremmen, 1999) were also computed, although because our items were counterbalanced across lists, this test is probably too conservative (see Raaijmakers, 2003, for discussion). This analysis yielded similar results to the linear mixed effects model. The main discrepancy was obtained in the English auxiliaries experiment in the *verb+1* region, where the interaction between grammaticality and attractor number didn't reach significance in the ANOVAs, although the qualitative pattern was in the expected direction. Furthermore, when the *verb+1* and *verb+2* regions are pooled for analysis, the interaction becomes significant by-subjects ($F_{1,31} = 5.32, p < .05$) and marginal by-items ($F_{2,47} = 2.96, p = .09$).

Spanish main verbs	By participants			By items			MinF'		
	df	F1	p	df	F2	p	df	minF'	p
<i>Verb region</i>									
Grammaticality	1,31	1.98	0.17	1,47	2.76	0.10	1,68	1.15	0.29
Attractor number	1,31	0.09	0.77	1,47	0.03	0.86	1,73	0.02	0.88
Gram × Attr number	1,31	0.42	0.52	1,47	0.41	0.52	1,75	0.21	0.65
<i>Verb+1 region</i>									
Grammaticality	1,31	62.62	<.01	1,47	106.33	<.01	1,64	39.41	<.01
Attractor number	1,31	9.40	<.01	1,47	6.12	<.05	1,78	3.71	<.07
Gram × Attr number	1,31	6.76	<.05	1,47	4.48	<.05	1,78	2.70	0.10
Spanish auxiliaries									
<i>Verb region</i>									
Grammaticality	1,31	0.05	0.82	1,47	0.00	0.95	1,54	0.00	0.95
Attractor number	1,31	0.57	0.46	1,47	0.18	0.67	1,71	0.14	0.71
Gram × Attr number	1,31	1.40	0.25	1,47	0.94	0.34	1,78	0.56	0.46
<i>Verb+1 region</i>									
Grammaticality	1,31	49.40	<.01	1,47	98.10	<.01	1,60	32.86	<.01
Attractor number	1,31	0.19	0.67	1,47	0.29	0.59	1,66	0.11	0.74
Gram × Attr number	1,31	10.13	<.01	1,47	11.18	<.01	1,73	5.31	<.05
<i>Verb+2 region</i>									
Grammaticality	1,31	38.35	<.01	1,47	45.49	<.01	1,72	20.81	<.01
Attractor number	1,31	2.90	0.10	1,47	1.47	0.23	1,77	0.97	0.33
Gram × Attr number	1,31	0.19	0.66	1,47	0.07	0.80	1,72	0.05	0.83
English auxiliaries									
<i>Verb region</i>									
Grammaticality	1,31	0.13	0.72	1,47	0.03	0.86	1,66	0.03	0.87
Attractor number	1,31	4.37	<.05	1,47	6.41	<.05	1,67	2.60	0.11
Gram × Attr number	1,31	0.66	0.42	1,47	0.32	0.57	1,77	0.22	0.64
<i>Verb+1 region</i>									
Grammaticality	1,31	20.45	<.01	1,47	18.89	<.01	1,76	9.82	<.01
Attractor number	1,31	6.03	<.05	1,47	5.10	<.05	1,77	2.76	0.10
Gram × Attr number	1,31	2.31	0.14	1,47	3.16	<.09	1,69	1.34	0.25
<i>Verb+2 region</i>									
Grammaticality	1,31	12.74	<.01	1,47	2.87	0.10	1,66	2.34	0.13
Attractor number	1,31	5.65	<.05	1,47	6.41	<.05	1,73	3.00	<.09
Gram × Attr number	1,31	2.90	0.10	1,47	1.69	0.20	1,78	1.07	0.30

Table S1. Omnibus repeated measures analysis of variance for the three experiments reported in this paper. Significant effects are in bold font, marginal effects with $.05 < p < .1$ are in bold and italic font.

Appendix C. Item sets for Experiments 4 & 5

1. The housemaids encountered the lord/serf/Eric/Bert in the local park. They laughed at his jokes on the way home.
2. The girls asked the lord/serf/Eric/Bert for the latest gossip. They gasped at his sordid stories about the aristocracy.
3. The doormen ushered the student/vagrant/Wesley/Kellen away from the building. They pointed to his bag, which lay beside the door.

4. The girls saw the student/vagrant/Wesley/Kellen in the street in the city center. They talked about his clothes, which were rather odd.
5. The constables suspected the criminal/arsonist/George/Grover of the break-in. They searched in his bag for evidence of the crime.
6. The old couple begged the criminal/arsonist/George/Grover for mercy. They appealed to his pity and religious values.
7. The neighbors visited the professor/navigator/Michael/Landon a couple of days ago. They spoke about his family most of the time.
8. The dogs greeted the professor/navigator/Michael/Landon at the front door. They jumped on his clothes leaving a muddy patch.
9. The audience applauded the artist/oboist/Brian/Irvin for a magnificent performance. They clapped at his entourage appreciatively, also.
10. The ladies watched the artist/oboist/Brian/Irvin yesterday afternoon. They stared at his hat, which was very outrageous.
11. The girls met the commander/aristocrat/Anthony/Wilhelm during the party. They heard about his wealth later in the evening.
12. The servants welcomed the commander/aristocrat/Anthony/Wilhelm back to the mansion. They marveled about his fiancée from the distant land of India.
13. The understudies envied the actor/tenor/Jack/Alva for being so successful. They longed for his wealth and fame all the time.
14. The crowd thrilled the actor/tenor/Jack/Alva with a standing ovation. They responded to his performance in an emotional way.
15. The ministers blamed the king/czar/Ryan/Ivor for the crisis in the country. They asked for his resignation a couple of days ago.
16. The people respected the king/czar/Ryan/Ivor for having served the country in many battles. They flocked to his funeral, which was held in the capital.
17. The mothers told the farmer/cobbler/Daniel/Wilbur about the suffocating smoke. They complained about his bonfire, which caused a lot of pollution.
18. The brothers accompanied the farmer/cobbler/Daniel/Wilbur to the local market. They listened to his stories about the little town.

19. The reporters asked the doctor/envoy/Tyler/Seely about the incident. They gaped at his clothes that were ripped to pieces.
20. The assistants disliked the doctor/envoy/Tyler/Seely in recent months. They grumbled about his rudeness and bad behavior.
21. The warriors fought the soldier/samurai/William/Kenton for many hours. They succumbed to his overpowering strength after a long struggle.
22. The peasants feared the soldier/samurai/William/Kenton during the drawn-out war. They worried about his threats all the time.
23. The advisors blamed the president/emissary/Matthew/Elwood for the recent problems. They disagreed with his plans to negotiate a new deal.
24. The ambassadors asked the president/emissary/Matthew/Elwood for more time. They thought about his proposal for many hours.
25. The villagers worshipped the priest/druid/Peter/Silas in a very literal sense. They prayed for his well-being every morning.
26. The women asked the priest/druid/Peter/Silas for some advice. They read from his book for an answer to their dilemma.
27. The townspeople ignored the stranger/bohemian/Jeremy/Horace almost entirely. They disapproved of his different habits and customs.
28. The girls greeted the stranger/bohemian/Jeremy/Horace with open arms. They had waited for his arrival for a very long time.
29. The family loved the lawyer/mogul/John/Wiley loyally and steadfastly. They mourned over his death for many weeks.
30. The ladies recognized the lawyer/mogul/John/Wiley at the special ball. They pointed to his fashionable suit with admiration.
31. The visitors awaited the director/hotelier/Andrew/Hubert for at least half an hour. They stood in his reception area talking very loudly.
32. The mobsters ushered the director/hotelier/Andrew/Hubert back into the house. They gestured to his children in a menacing way.
33. The women obeyed the policeman/legionary/Cameron/Willard in shocked confusion. They shuddered at his commands, which were sharp and harsh.

34. The crowd disliked the policeman/legionary/Cameron/Willard very intensely. They responded to his request in a very angry way.
35. The journalists interviewed the congressman/legislator/Patrick/Norbert about the latest scandal. They inquired about his extramarital affairs from the past.
36. The ministers condemned the congressman/legislator/Patrick/Norbert for the diplomatic disaster. They disapproved of his speech in the United Nations meeting.
37. The housekeepers called the plumber/roofer/Alex/Elmer to work on the leak. They waited for his arrival anxiously at the gate.
38. The housewives liked the plumber/roofer/Alex/Elmer from the house next door. They asked for his phone number at the town meeting.
39. The agents arrested the terrorist/racketeer/Joseph/Alonzo yesterday at the public library. They took away his passport immediately after the arrest.
40. The historians described the terrorist/racketeer/Joseph/Alonzo as a fearless character. They wrote about his childhood experience in the Second World War.
41. The employees criticized the executive/overseer/Richard/Emmett during the union meeting. They disapproved of his oppressive demeanor, which was completely unacceptable.
42. The clients thanked the executive/overseer/Richard/Emmett for organizing the event. They identified with his enthusiasm for improving the company's image.
43. The fans applauded the player/cellist/Scott/Boyce at the end of the event. They asked for his autograph frantically in the parking lot afterwards.
44. The audience gave the player/cellist/Scott/Boyce a standing ovation. They waited for his appearance at the exit with excitement.
45. The readers boycotted the reporter/columnist/Robert/Edmund for the recent controversial piece. They called for his immediate resignation from the editorial board.
46. The publishers hired the reporter/columnist/Robert/Edmund without any hesitation. They believed in his ability to attract new readers.
47. The patients trusted the therapist/allergist/Carlos/Colton despite the recent rumors. They swore by his qualifications in front of the press.
48. The interns offended the therapist/allergist/Carlos/Colton by refusing to help. They wore on his patience without realizing it.

49. The judges disqualified the athlete/hurdler/Aaron/Lonnie before the competition started. They had heard about his drug abuse history the previous week.
50. The paramedics carried the athlete/hurdler/Aaron/Lonnie off the running track. They looked after his wound at the paramedic station in the corner.
51. The residents blamed the sheriff/rancher/Jordan/Ramsay for the recent fire. They talked about his negligence constantly around the town.
52. The dogs scared the sheriff/rancher/Jordan/Ramsay during the visit to the farm. They barked at his horse until it left the property.
53. The secretaries sued the psychiatrist/orthodontist/Christian/Cornelius after years of verbal abuse. They complained about his horrible temper and violent behavior.
54. The statements embarrassed the psychiatrist/orthodontist/Christian/Cornelius during the malpractice trial. They focused on his longstanding addiction to prescription sleeping pills.
55. The ambassadors met the emperor/monarch/Thomas/Reuben during the country's best years. They wrote about his charisma and cleverness using vivid examples.
56. The senators hated the emperor/monarch/Thomas/Reuben after the news of the scandal hit. They scoffed at his requests for support without even thinking.
57. The critics praised the photographer/screenwriter/Nicholas/Wendell in all the city's newspapers. They raved about his skillful portrayal of modern living in Asia.
58. The reviews reached the photographer/screenwriter/Nicholas/Wendell only after several days. They came to his attention because of a spiteful blogger.
59. The servants robbed the knight/regent/Adam/Ollie of the expensive paintings. They fled from his mansion immediately to find a hiding place.
60. The princesses avoided the knight/regent/Adam/Ollie because of the recent rumors. They saw through his attempts to appear virtuous before the court.
61. The organizers offended the scientist/geographer/Timothy/Everett by requesting the original data. They disagreed with his arguments in favor of the existing views.
62. The students bothered the scientist/geographer/Timothy/Everett with simple-minded questions after class. They stayed for his office hours to ask about the exam.
63. The squirrels delighted the runner/cyclist/Mark/Cecil by scurrying around the tree. They lightened up his mood with their endless playful frolicking.

64. The motorists followed the runner/cyclist/Mark/Cecil along the residential street. They interfered with his workout by honking loudly and aggressively.
65. The prisoners bribed the officer/jailer/Jason/Hiram to gain access to the tobacco. They relied on his staff to complete the transactions discreetly.
66. The blackmailers threatened the officer/jailer/Jason/Hiram in a terrifying letter. They hinted at his hidden criminal past and recent indiscretions.
67. The defendants urged the judge/jurist/Sean/Percy to reconsider the evidence. They pointed to his reputation of being a truth seeker.
68. The victims thanked the judge/jurist/Sean/Percy for doing justice. They spoke of his impartiality in front of the press.
69. The actors called the producer/dramatist/Justin/Elbert to discuss the script. They asked for his help to resolve some issues with the dialog.
70. The actors met the producer/dramatist/Justin/Elbert at the cocktail party. They inquired about his latest play and the possibility of getting an audition.
71. The foxes attacked the hunter/poacher/Paul/Clyde yesterday during the hunting trip in the national park. They chewed on his rifle and emergency communication device.
72. The protesters urged the hunter/poacher/Paul/Clyde to stop killing wild animals. They pointed to his cruelty and handed over a petition.
73. The journalists interviewed the coach/goalie/Kevin/Rufus during the playoffs. They asked about his confidence in the team's performance.
74. The audience booed the coach/goalie/Kevin/Rufus for the historic loss. They jeered at his overall incompetence during the last few games.
75. The pirates released the captain/mariner/Martin/Roscoe after receiving the ransom. They laughed about his fearfulness long after the release.
76. The villagers warned the captain/mariner/Martin/Roscoe about the upcoming storm. They worried about his safety out alone in the open sea.
77. The gallery owners called the painter/artisan/James/Vernon to report on the sales. They asked for his permission to lower the price on some items.
78. The onlookers congratulated the painter/artisan/James/Vernon for the originality of the work. They marveled at his creativity and use of innovative techniques.

79. The clients visited the attorney/paralegal/David/Elijah at the law firm in the morning. They asked about his progress preparing for the case.
80. The customers contacted the attorney/paralegal/David/Elijah to discuss the bills. They complained about his exorbitant fees and poor service.

Appendix D. Item sets for Experiments 6 & 7

German materials (Experiment 6)

The experiment contained 64 sets of 8 conditions.

Phonological conditions

1. Die Nachbarinnen mochten den **Zeichenlehrer/Administrator**, der im obersten Stockwerk wohnte. Sie gingen sicher, dass seine/die **Zeitung** nicht aus seinem Briefkasten geklaut wurde.
2. Die Kundenberaterinnen beruhigten den **Konditor/Algerier**, dessen Kreditkarte gestohlen worden war. Sie sagten, dass sein/das **Konto** gesperrt worden sei und er sich keine Sorgen machen müsse.
3. Die Kolleginnen bemitleideten den **Optiker/Pendler**, der an einem entzündeten Blinddarm litt. Sie wussten, dass seine/die **Operation** direkt bevorstand und wollten ihn beruhigen.
4. Die Tiertrainerinnen hatten Mitleid mit dem **Wärter/Zoologen**, der starke Rückenschmerzen hatte. Sie hofften, dass seine/die **Wärmflasche** die Beschwerden etwas lindern würde.
5. Die Mädchen riefen den **Wachmann/Dompteur**, als ein wilder Hund angerannt kam. Sie hofften, dass seine/die **Waffen**, die er immer bei sich trug, sie beschützen würden.
6. Mehrere Zuhörerinnen sprachen den **Redner/Piloten** an, der die Veranstaltung verließ. Sie sahen, dass sein/die **Regenschirm**, den er unter dem Arm trug, kaputt war.
7. Die Organisatorinnen bat den **Autor/Leiter**, pünktlich zu sein. Sie wussten, dass sein/das **Auto**, das er erst vor kurzem gekauft hatte, kaputt war und er ein Taxi gerufen hatte.
8. Die Ärztinnen berieten den **Kontrolleur/Auktionator**, der gesünder leben wollte. Sie bemängelten, dass seine/die **Kondition**, die er bei dem Fitnesstest gezeigt hatte, sehr schlecht war.
9. Die Kassiererinnen riefen nach dem **Fleischer/Kapellmeister**, der schon dabei war zu gehen. Sie wussten, dass seine/die **Flasche Wein**, die er hatte zurücklegen lassen, noch an der Kasse stand.

10. Die Frauen auf dem Laufband drehten sich nach dem **Handwerker/Komponisten** um, als ein lautes Krachen zu hören war. Sie hofften, dass seine/die **Hanteln**, mit denen er trainierte, nicht wieder auf den Boden gefallen war.
11. Die Mieterinnen hörten den **Hausmeister/Metallarbeiter** fluchen, als das Treppenhaus gereinigt wurde. Sie befürchteten, dass seine/die **Haut**, die nicht durch Handschuhe geschützt war, durch das ätzende Putzmittel verletzt worden war.
12. Die Inspektorinnen verärgerten den **Bauern/Franzosen** bei der Besichtigung des Hofs. Sie sagten, dass sein/der **Baum**, der im Vorgarten stand, gefällt werden müsste, weil die Wurzeln den Bürgersteig beschädigten.
13. Die Dozentinnen warnten den **Studenten/Mitarbeiter**, der versuchte bei der Klausur abzuschreiben. Sie hatten bemerkt, dass sein/der **Stuhl** viel zu nah an seinem Banknachbarn stand.
14. Die Frauen grüßten den **Hufschmied/Juniorchef** im Vorbeigehen. Sie riefen, dass sein/der **Hut**, den er neu gekauft hatte, ihm sehr gut stand.
15. Die Umzugshelferinnen verhandelten mit dem **Kellner/Mieter** bevor der Umzug begann. Sie hatten gesehen, dass sein/der **Keller** mit Kisten vollgestellt war und verlangten mehr Geld.
16. Die Kundinnen dankten dem **Mechaniker/Verwalter** am Telefon. Sie bestätigten, dass seine/die **Meldung** sehr wichtig war und es richtig gewesen war, sie sofort anzurufen.
17. Die Freundinnen kannten den **Kritiker/Theologen** sehr gut. Sie wussten, dass seine/die **Krise**, die er gerade durchmachte, nicht leicht zu überwinden sein würde.
18. Die Kolleginnen umarmten den **Trainer/Forscher**, als das Spiel zu Ende war. Sie sahen, dass seine/die **Tränen**, die die Wangen herunterflossen, ehrlich waren.
19. Die Kellnerinnen sprachen mit dem **Tierarzt/Apotheker**, der auf die Bestellung wartete. Sie versicherten, dass sein/das **Tiramisu** bald kommen würde und fragten, ob er noch etwas anderes wünschte.
20. Die Köchinnen konnten den **Schwimmer/Taxifahrer** gut leiden, der häufig zum Mittagessen kam. Sie stellten sicher, dass sein/der **Schweinebraten**, den er so gerne mochte, stets bereit war, wenn er vorbeikam.
21. Die Zimmermädchen warnten den **Hotelier/Anfänger**, dass die Hotelleitung großen Wert auf ein gepflegtes Äußeres legte. Sie starnten auf seine/die **Hose**, die am Knie ein Loch hatte.
22. Die Managerinnen lobten den **Verkäufer/Erfinder**, als die Inventur beendet war. Sie versprachen, dass sein/der **Vertrag** definitiv verlängert werden würde.
23. Die Kundenberaterinnen telefonierten mit dem **Zahnarzt/Exporteur** wegen der Lieferung des neuen Behandlungsstuhls. Sie sagten, dass seine/die **Zahlung** noch nicht bei ihnen eingetroffen sei.

24. Die Krankenschwestern kümmerten sich um den **Schauspieler/Sozialisten** in der Klinik. Sie gingen sicher, dass sein/die **Schaumbad** die perfekte Temperatur erreicht hatte, bevor er in die Wanne stieg.
25. Die Fußgängerinnen starrten den **Fahrer/Spieler** an, als das Auto auf dem Parkplatz hielt. Sie dachten, dass seine/die **Fahne** an dem Wagen sicher die anderen Verkehrsteilnehmer ablenken würde.
26. Die Ärztinnen sprachen mit dem **Bäcker/Jockey** über die Schmerzen. Sie vermuteten, dass sein/das **Becken**, das ihm Probleme machte, sehr unter der täglichen Belastung bei der Arbeit gelitten hatte.
27. Die Krankenschwestern sorgten sich um den **Arzt/Freund**, der bei dem Autounfall verletzt worden war. Sie waren überzeugt, dass sein/der **Arm** dringend geröntgt werden müsse.
28. Die Hausfrauen halfen dem **Briefträger/Steurberater**, nach dem Unfall aufzustehen. Sie sagten, dass seine/die **Brille**, die zerbrochen war, wahrscheinlich repariert werden könnte.
29. Die Freundinnen holten den **Stadtrat/Ausländer** vom Bahnhof ab. Sie fragten, ob seine/die **Stadtrundfahrt** interessant gewesen sei.
30. Die Gastgeberinnen kritisierten den **Nachrichtensprecher/Geschichtsschreiber** nach dem Abendessen. Sie hatten gesehen, dass sein/der **Nachtisch** von ihm nicht einmal probiert worden war.
31. Die Vermieterinnen zeigten den **Bewohner/Fischer** an, nachdem mehrere Mängel gefunden worden waren. Sie sagten, dass sein/der **Bestechungsversuch** alles nur noch schlimmer gemacht hätte.
32. Die Nachbarinnen machten dem **Klempner/Bassisten** ein Kompliment bei der Hauseinweihungsfeier. Sie fanden, dass sein/das **Klettergerüst**, das mitten im Garten stand, viel schöner war als ihres.
33. Die Sekretärinnen rieten dem **Anwalt/Senator**, etwas anderes anzuziehen. Sie erklärten, dass sein/der **Anzug** eine zu aufdringliche Farbe habe und nicht angemessen sei.
34. Mehrere Besucherinnen beschwerten sich über den **Priester/Herzog** nach einer öffentlichen Rede. Sie fanden, dass sein/das **Prinzip**, welches in Bezug zur Armut zum Ausdruck kam, nicht richtig war.
35. Die Mädchen riefen hinter dem **Bibliothekar/Fernfahrer** her, der dabei war zu gehen. Sie hatten bemerkt, dass seine/die **Bibel**, die er immer bei sich hatte, auf einer Bank liegengeblieben war.
36. Die Bankkauffrauen telefonierten mit dem **Abteilungsleiter/Parteimitglied** wegen der finanziellen Schwierigkeiten. Sie bedauerten, dass seine/die **Absicht**, den Angestellten dieses Jahr einen Bonus zu zahlen, leider unmöglich war.
37. Die Sekretärinnen wunderten sich über den **Programmierer/Sprachforscher**, der sonst immer gute Arbeit leistete. Sie hörten, dass sein/das **Projekt**, das er bei der Konferenz vorstellen wollte, nicht gut durchdacht sei.

38. Die Kellnerinnen ignorierten den **Anglisten/Cellisten**, der häufig ins Café kam. Sie lachten über sein/das **Angebot**, mit ihnen nach Feierabend essen zu gehen.
39. Die Managerinnen widersprachen dem **Rechenkünstler/Anlageberater** während des Meetings. Sie sagten, dass seine/die **Rechtfertigung** des Budgets nicht schlüssig war.
40. Alle Aktivistinnen unterstützten den **Produzenten/Intellektuellen** aus voller Überzeugung. Sie wussten, dass seine/die **Protestaktion** gegen schlechte Arbeitsbedingungen in den Medien sehr wichtig war.
41. Die Investorinnen ärgerten sich sehr über den **Unternehmer/Nachfolger**, der einfach das Meeting verließ. Sie fanden, dass seine/die **Unterbrechung** der Verhandlung einen Mangel an Respekt zeigte.
42. Die Schülerinnen kritisierten den **Bauarbeiter/Landwirt**, der durch die Innenstadt ging. Sie waren entsetzt, dass sein/der **Bauch**, der unter dem Hemd hervor quoll, ständig zu sehen war.
43. Die Reporterinnen interviewten den **Verteidiger/Innenminister** nach der Urteilsverkündung. Sie hofften, dass sein/der **Versuch**, die Sachlage zu beschönigen, zu einem riesigen Medienspektakel führen würde.
44. Die Hausfrauen grüßten den **Fotografen/Gesandten**, der in der Wohnung im Erdgeschoss wohnte. Sie erklärten, dass seine/die **Formalität**, die er ihnen gegenüber zeigte, völlig unnötig sei.
45. Die Wählerinnen zwangen den **Kommunisten/Politiker**, der zum Minister ernannt worden war, zurückzutreten. Sie meinten, dass sein/der **Kommentar** zur Frauenquote zu sexistisch gewesen sei.
46. Die Sekretärinnen beruhigten den **Uhrmacher/Veteranen** am Telefon. Sie versprachen, dass seine/die **Urkunde** sogleich abgeschickt werden würde und bald bei ihm ankommen würde.
47. Die Verkäuferinnen berieten den **Kassierer/Methodisten** beim Kauf einer neuen Stereoanlage. Sie warnten, dass seine/die **Kassette**, die ihm so wichtig war, auf einem CD-Spieler nicht abgespielt werden könne.
48. Die Mitbewohnerinnen begannen, sich über den **Frisör/Barmann** zu ärgern, der erst vor kurzem eingezogen war. Sie fanden, dass seine/die **Fritteuse** in der Küche zu viel Platz einnahm.
49. Die Juristinnen verhörten den **Jäger/Häftling** während der Verhandlung. Sie wussten, dass sein/der **Jähzorn**, für den er bekannt war, wahrscheinlich zu dem verantwortungslosen Verhalten geführt hatte.
50. Die Krankenschwestern beruhigten den **Taucher/Pächter** nach der Visite. Sie sagten, dass seine/die **Tauglichkeit** für seinen Beruf durch die Operation nicht beeinträchtigt werden würde.
51. Die Sekretärinnen unterhielten sich mit dem **Akademiker/Bauherrn** an der Rezeption. Sie sahen, dass seine/die **Aktentasche**, die er unter dem Arm trug, offen war und hofften, dass nichts herausfallen würde.

52. Die Studentinnen trafen den **Architekten/Historiker** im Büro. Sie fragten, ob sein/das **Archiv**, auf das nur er Zugriff hatte, möglicherweise hilfreiche Dokumente für ihr Projekt beinhaltete.
53. Mehrere Journalistinnen interviewten den **Richter/Kanzler** nach der Pressekonferenz. Sie fragten, ob sein/der **Ring** an der linken Hand irgendeine wichtige Bedeutung habe.
54. Die Stewardessen lächelten den **Reiter/Sportler** an, der ins Flugzeug einstieg. Sie fragten, ob seine/die **Reise** zu der Veranstaltung angenehm gewesen sei.
55. Die Filmstudentinnen befragten den **Regisseur/Experten** zu dem Film. Sie fanden, dass seine/die **Reaktion** auf das Ende zu emotional gewesen sei.
56. Die Personalmanagerinnen kritisierten den **Statisten/Ökonom**, der Probleme hatte, Anweisungen zu folgen. Sie sagten, dass sein/der **Starrsinn**, den er von Anfang an gezeigt hatte, seiner weiteren Karriere sehr im Weg stehen würde.
57. Die Empfangsdamen gaben sich Mühe, den **Komponisten/Besitzer** beim Theaterbesuch zufriedenzustellen. Sie gingen sicher, dass sein/der **Kollege**, auf den er wartete, einen guten Platz bekam.
58. Die Putzfrauen suchten den **Matrosen/Raucher** auf dem ganzen Deck. Sie fanden, dass seine/die **Matratze** gewendet werden müsse und brauchten dafür seine Hilfe.
59. Die Verwalterinnen beschwerten sich beim **Inhaber/Helper** am Ende des Monats. Sie sagten, dass seine/die **Investition** in das Projekt nicht groß genug gewesen sei.
60. Die Beraterinnen riefen den **Rentner/Gelehrten** nachmittags zu Hause an. Sie versprachen, dass seine/die **Rendite** dieses Jahr besonders gut sein würde.
61. Die Sekretärinnen rieten dem **Vermieter/Chauffeur**, die Bank zu kontaktieren. Sie bestätigten, dass seine/die **Versicherung** den Schaden übernehmen würde.
62. Die Buchhalterinnen bemitleideten den **Redakteur/Manager**, der bei dem Sturm rausgehen musste. Sie hofften, dass sein/der **Regenschirm** nicht kaputt gehen würde.
63. Die Soldatinnen wunderten sich über den **Trommler/Söldner**, der ein nasses Hemd trug. Sie wussten nicht, dass sein/der **Trockner**, den er erst vor kurzem gekauft hatte, schon kaputt war.
64. Die Ärztinnen rieten dem **Karikaturenzeichner/ Langstreckenläufer**, eine Auszeit zu nehmen. Sie erklärten, dass sein/das **Kardiogramm**, das endlich ausgewertet worden war, ihnen Sorgen machte.

Semantic conditions

1. Die Nachbarinnen mochten den **Zeichenlehrer/ Administrator**, der im obersten Stockwerk wohnte. Sie fanden, dass sein/das **Bild**, an dem er in seiner Freizeit gearbeitet hatte und das jetzt im Hausflur hing, sehr gut geworden war.

2. Die Hausfrauen lobten den **Konditor/Algerier**, der ein neues Geschäft in der Nachbarschaft eröffnete. Sie meinten, dass sein/das **Gebäck**, das er verkaufte, sehr gut war.
3. Die Kolleginnen bemitleideten den **Optiker/Pendler**, der an einem entzündeten Blinddarm litt. Sie hatten gehört, dass seine/die **Brille**, ohne die er nichts sehen konnte, im Krankenhaus verloren gegangen war.
4. Die Tiertrainerinnen baten den **Wärter/Zoologen**, mehr auf die Sicherheit des Zoos zu achten. Sie sahen, dass seine/die **Schlüssel**, die er sonst am Gürtel trug, heruntergefallen waren und wiesen ihn darauf hin.
5. Die Mädchen bewunderten den **Wachmann/Dompteur**, der makellos aussah. Sie fanden, dass seine/die **Uniform**, die er zur Arbeit trug, ihm sehr gut stand.
6. Mehrere Zuhörerinnen sprachen den **Redner/Piloten** nach der Pressekonferenz an. Sie fragten, ob seine/die **Notizen**, die er sich gemacht hatte, online veröffentlicht würden.
7. Die Verlegerinnen unterhielten sich mit dem **Autor/Leiter**, der etwas veröffentlichen wollte. Sie fragten, ob sein/das **Manuskript**, an dem er immer noch arbeitete, bald fertig sein würde.
8. Die Besitzerinnen trafen den **Kontrolleur/Auktionator**, der die Ware begutachten wollte. Sie bat, dass seine/die **Formulare**, die er noch vollständig ausfüllen musste, ihnen bald zugesandt würden.
9. Die Kassiererinnen riefen nach dem **Fleischer/Kapellmeister**, der schon am Gehen war. Sie wussten, dass sein/der **Schinken**, den er hatte zurücklegen lassen, noch an der Kasse lag.
10. Die Produktionsassistentinnen waren gleich zur Stelle, nachdem der **Handwerker/Komponist** seine Arbeit beendet hatte. Sie sahen, dass seine/die **Säge**, die er zum Aufbau der Bühne benötigt hatte, leider kaputt gegangen war.
11. Die Mieterinnen baten den **Hausmeister/Metallarbeiter** nach Vollendung des Umbaus zu putzen. Sie sagten, dass sein/der **Besen** in der Abstellkammer zu finden sei.
12. Die Inspektorinnen verärgerten den **Bauern/Franzosen** bei der Besichtigung des Hofs. Sie kritisierten, dass sein/der **Traktor**, der noch ganz neu war, nicht den Vorschriften entsprach.
13. Die Sekretärinnen baten den **Studenten/Mitarbeiter**, die Unterlagen rechtzeitig zu bearbeiten. Sie hofften, dass sein/der **Professor**, der das Projekt leitete, dann zufrieden sein würde.
14. Die Frauen grüßten den **Hufschmied/Juniorchef** im Vorbeigehen. Sie bemerkten, dass sein/das **Pferd**, auf dem er häufig ausritt, etwas nervös wirkte.
15. Die Besitzerinnen verhandelten mit dem **Kellner/Mieter**, bevor der Mietvertrag unterschrieben wurde. Sie stellten sicher, dass sein/das **Restaurant**, das hoch verschuldet war, ihn nicht am Zahlen hindern würde.

16. Die Kundinnen dankten dem **Mechaniker/Verwalter** am Telefon. Sie sagten, dass sein/das **Auto**, das er ihnen geliehen hatte, sehr gut fuhr.
17. Die Journalistinnen kannten den **Kritiker/Theologen** sehr gut. Sie erwarteten, dass seine/die **Rezension**, die er vor kurzem geschrieben hatte, zu interessanten Diskussionen führen würde.
18. Die Kolleginnen machten sich über den **Trainer/Forscher** lustig, als das Spiel endete. Sie sahen, dass seine/die **Sportschuhe**, auf die er so stolz war, mit Schlamm überzogen waren.
19. Die Spaziergängerinnen sprachen mit dem **Tierarzt/Apotheker**, der immer im Park spazieren ging. Sie versicherten, dass sein/der **Hund**, der immer mit dabei war, ein äußerst schönes Tier sei.
20. Die Mädchen redeten gerne mit dem **Schwimmer/Taxifahrer**, der immer in das Sportstudio gegenüber kam. Sie sagten, dass seine/die **Badehose**, die jedes Mal seinen Körper zierte, ein schönes Muster hatte.
21. Die Besucherinnen gaben dem **Hotelier/Anfänger** eine schlechte Rückmeldung. Sie beschwerten sich, dass seine/die **Rezeption** schmutzig und mit Papier vollgestopft war.
22. Die Managerinnen dankten dem **Verkäufer/Erfinder**, der den Kundenservice verbessert hatte. Sie sagten, dass seine/die **Kasse**, an der er lange gearbeitet hatte, allen viel Zeit gespart hatte.
23. Die Kolleginnen berieten den **Zahnarzt/Exporteur**, der an starken Kopfschmerzen litt. Sie sagten, dass seine/die **Weisheitszähne**, die er vor kurzem bekommen hatte, sich wahrscheinlich entzündet hatten.
24. Die Assistentinnen benachrichtigten den **Schauspieler/Sozialisten**, als alles bereit war. Sie sagten, dass sein/der **Monolog**, den er so lange geübt hatte, bestimmt gut ankommen würde.
25. Die Polizistinnen hielten den **Fahrer/Spieler** am Zebrastreifen an. Sie kritisierten, dass sein/das **Lenkrad** und die Rückscheinwerfer in einem sehr schlechten Zustand waren.
26. Die Vermieterinnen sprachen mit dem **Bäcker/Jockey** über die Beschwerden der anderen. Sie sagten, dass sein/der **Ofen**, wenn zu heiß aufgedreht, die Hauswände erwärme und die Nachbarn zu sehr ins Schwitzen bringe.
27. Die Seniorinnen mochten den **Arzt/Freund** aufgrund seines Wissens. Sie wussten, dass seine/die **Untersuchung**, die immer mit größter Genauigkeit durchgeführt wurde, stets zu einem guten Ergebnis führte.
28. Die Hausfrauen grüßten den **Briefträger/Steuerberater** jeden Morgen. Sie hofften, dass sein/das **Paket**, das er unter dem Arm trug, für sie war.
29. Die Moderatorinnen sprachen mit dem **Stadtrat/Ausländer** über Immigrationspolitik. Sie sagten, dass seine/die **Pressekonferenz**, bei der er menschliches Miteinander einforderte, die Menschen sehr berührt habe.

30. Die Freundinnen kritisierten den **Nachrichtensprecher/Geschichtsschreiber** nach dem Abendessen. Sie dachten, dass seine/die **Ansage**, dass er sich an eine strenge Diät halten wolle, nicht ernst zu nehmen sei.
31. Die Inspektorinnen zeigten den **Bewohner/Fischer** an, nachdem mehrere Mängel gefunden worden waren. Sie fanden, dass sein/das **Haus** in einem sehr schlechten Zustand war.
32. Die Nachbarinnen lobten den **Klempner/Bassisten** bei der Hauseinweihungsfeier. Sie fanden, dass seine/die **Rohre** in allen Zimmern sehr gut verputzt worden waren und man sie kaum sehen konnte.
33. Die Moderatorinnen ließen den **Anwalt/Senator** nicht zu Ende sprechen. Sie erklärten, dass sein/der **Einspruch** nicht berechtigt sei.
34. Mehrere Besucherinnen beschwerten sich über den **Priester/Herzog** nach einer öffentlichen Rede. Sie fanden, dass seine/die **Kirchengemeinde** solche Ansichten nicht dulden sollte.
35. Die Mädchen riefen hinter dem **Bibliothekar/Fernfahrer** her, der dabei war zu gehen. Sie hatten bemerkt, dass sein/das **Buch**, in das er bis eben so vertieft gewesen war, auf einer Bank liegengeblieben war.
36. Die Bankkauffrauen telefonierten mit dem **Abteilungsleiter/Parteimitglied** wegen der finanziellen Schwierigkeiten. Sie fragten, ob seine/die **Angestellten**, die schon lange für ihn arbeiteten, eine Lohnkürzung akzeptieren würden.
37. Die Sekretärinnen wunderten sich über den **Programmierer/Sprachforscher**, der sonst immer gute Arbeit leistete. Sie wussten nicht, dass sein/der **Computer**, auf den er für seine Arbeit angewiesen war, kaputt war.
38. Die Kellnerinnen ignorierten den **Anglisten/Cellisten**, der häufig ins Café kam. Sie fanden, dass seine/die **Übersetzung** der Speisekarte ins Englische, mit der er sie beeindrucken wollte, lächerlich war.
39. Die Professorinnen widersprachen dem **Rechenkünstler/Anlageberater** in Bezug auf seine Darlegung. Sie behaupteten, dass sein/der **Taschenrechner** eine Fehlfunktion haben müsse.
40. Alle Aktivistinnen unterstützten den **Produzenten/Intellektuellen** aus voller Überzeugung. Sie meinten, dass sein/der **Film**, an dem er mit vielen anderen zusammen gearbeitet hatte, eine wichtige Botschaft hatte.
41. Die Investorinnen ärgerten sich sehr über den **Unternehmer/Nachfolger**, der bei dem Meeting vorgestellt wurde. Sie dachten, dass sein/der **Geschäftsplan**, der vorgeschlagen wurde, schlecht durchdacht war.
42. Die Feriengäste beschwerten sich bei dem **Bauarbeiter/Landwirt** wegen des Hauses. Sie sagten, dass sein/das **Gerüst**, das vor kurzem aufgestellt worden war, die Aussicht ruinierte.
43. Die Reporterinnen interviewten den **Verteidiger/Innenminister** nach dem Skandal. Sie berichteten, dass sein/das **Plädoyer** gegen die Entscheidung für sehr viel Aufsehen gesorgt hatte.

44. Die Hausfrauen begrüßten den **Fotografen/Gesandten**, der gerade aus den Ferien zurückgekommen war. Sie fragten, ob seine/die **Kamera**, die er neu gekauft hatte, gut funktioniert hatte und wollten Bilder sehen.
45. Die Wählerinnen kritisierten den **Kommunisten/Politiker**, der zum Minister ernannt worden war. Sie meinten, dass sein/das **Stalinporträt** in seinem Büro als Wandschmuck unangebracht sei.
46. Die Sekretärinnen beruhigten den **Uhrmacher/Veteranen** am Telefon. Sie versprachen, dass sein/das **Vergrößerungsglas**, das er extra bestellt hatte, bald geliefert werde würde.
47. Die Verkäuferinnen berieten den **Kassierer/Methodisten** beim Kauf einer neuen Stereoanlage. Sie garantierten, dass sein/das **Geld** für das beste Modell reichen würde.
48. Die Mitbewohnerinnen begannen, sich über den **Frisör/Barmann** zu ärgern, der erst vor kurzem eingezogen war. Sie fanden, dass sein/der **Haarschnitt**, von dem er so begeistert war, ihn arrogant aussehen ließ.
49. Die Juristinnen verhörten den **Jäger/Häftling** während der Verhandlung. Sie fragten, ob sein/das **Gewehr** während des Vorfalls geladen gewesen sei.
50. Die Reiseleiterinnen redeten mit dem **Taucher/Pächter** während der Exkursion. Sie versprachen, dass sein/der **Schnorchel**, den er benutzte, geprüft und sicher sei.
51. Die Sekretärinnen unterhielten sich mit dem **Akademiker/Bauherrn** als das Telefon klingelte. Sie sagten, dass sein/der **Vortrag**, den er gestern gehalten hatte, sehr gelobt worden war.
52. Die Studentinnen besuchten den **Architekten/Historiker** bei der Arbeit. Sie hatten gehört, dass sein/das **Gebäude**, in dem er ein Büro hatte, unter Denkmalschutz stand und sehr interessant war.
53. Mehrere Journalistinnen interviewten den **Richter/Kanzler** nach der Pressekonferenz. Sie unterstellten, dass sein/das **Urteil** in der Angelegenheit durch seine eigenen Interessen kompromittiert war.
54. Die Reporterinnen interviewten den **Reiter/Sportler** nach dem Rennen. Sie bewunderten seinen/den **Sattel**, den er bei dem Derby benutzt hatte, weil er so außergewöhnlich war.
55. Die Studentinnen befragten den **Regisseur/Experten** hinsichtlich der Casting-Auswahl. Sie bezweifelten, dass sein/der **Kameramann** die beste Wahl für den Film gewesen sei.
56. Die Personalmanagerinnen kritisierten den **Statisten/Ökonom**, der Probleme hatte, Anweisungen zu folgen. Sie warnten, dass sein/der **Auftritt** am vorherigen Tag ihn möglicherweise sogar seinen Job kosten könnte.
57. Die Putzfrauen kritisierten den **Komponisten/Besitzer** nach der Reinigung des Theatersaals. Sie sagten, dass seine/die **Noten**, mit denen er gestern geübt hatte, überall herumlagen.

58. Die Freundinnen luden den **Matrosen/Raucher** ein auf die kleine Insel mitzukommen. Sie sagten, dass sein/das **Boot**, das er eigentlich verkaufen wollte, sie dorthin bringen könnte.
59. Die Frauen mochten den **Inhaber/Helfer**, der beim Tragen der Einkaufstüten half. Sie besuchten seinen/den **Laden**, in dem er sich immer aufhielt, um sich zu bedanken.
60. Die Nachbarinnen trafen den **Rentner/Gelehrten** auf dem Weg nach Hause. Sie sagten, dass seine/die **Enkelkinder**, die mit ihm spazieren gingen, sehr wohlerzogen waren.
61. Die Sekretärinnen rieten dem **Vermieter/Chauffeur**, die Bank zu kontaktieren. Sie sagten, dass seine/die **Wohnung**, auf die er einen Kredit aufnehmen wollte, erst besichtigt werden müsste.
62. Die Kolleginnen mochten den **Redakteur/Manager** für den Sportbereich. Sie fanden, dass seine/die **Zeitschrift**, die er betreute, sehr lesenswert war.
63. Die Soldatinnen bezahlten den **Trommler/Söldner**, an der Parade teilzunehmen. Sie hofften, dass seine/die **Pauke**, die er mitbringen wollte, die Menge aufheitern würde.
64. Die Organisatorinnen bat den **Karikaturenzeichner/Langstreckenläufer**, die Siegerkunde auszufüllen. Sie bemerkten, dass sein/der **Stift**, den er aus seiner Tasche zog, schmierte.

English materials (Experiment 7)

The experiment contained 64 sets of 8 conditions.

Phonological conditions

1. Several publishers approached the **journalist/congressman** who had recently made headlines. They said that his/the **journey** into fame had only just begun.
2. The maids lectured the **carpenter/journalist** for being terribly messy. They stared at his/the **carpet**, which was covered with mud stains.
3. The military prosecutors questioned the **civilian/therapist** during the trial. They focused on his/the **theft** of confidential patient files.
4. Many people pitied the **photographer/therapist** who was afraid of spider. They knew that his/the **phobia** could be crippling.
5. Some housewives visited the **minister/hunter**, who was gardening in the front yard. They commented that his/the **mint** was coming in rather nicely.
6. The accountants met with the **reverend/minister** for several hours. They thoroughly examined his/the **revenue** from all of the investments he made last year.
7. The guests thanked the **manager/player** at the exclusive cocktail party. They walked around his/the **mansion** and praised the décor.

8. The receptionists flirted with the **attorney/manager** in the office. They delighted in his/the **attention**, and so their flirting began to get a little excessive.
9. The CIA agents scrutinized the **chairman/terrorist** for the suspicious transactions. They feared that his/the **charity** was merely a front for a more nefarious plot.
10. The TV producers insulted the **reporter/champion** who had become famous during the Olympics. They knew that his/the **reply** would be quick and insulting.
11. The actresses criticized the **chauffeur/bodyguard** for having a bad attitude. They couldn't stand his/the **chauvinism** because it had started to spread to the other men.
12. The agents were after the **sheriff/killer** for the recent murder. They said that his/the **shed** was full of incriminating evidence.
13. The hostage negotiators freed the **captain/doctor** with clever trickery. They reasoned with his/the **captors** until he finally was released.
14. Some robbers attacked a **knight/priest** who was traveling to Avignon. They struck his/the **knife** out of the palm of his hand and demanded all his money.
15. The girls thanked the **officer/agent** politely for the invitation. However, they were skeptical of his/the **offer** to dine at his home that evening.
16. The plaintiffs argued with the **judge/lawyer** after the trial. They thought that his/the **justification** of the defendant's actions was completely inappropriate.
17. Many critics admired the **producer/graduate** from the acclaimed art school. They referred to his/the **program** as the best one in town.
18. The rebels deposed the **general/president** who had recently risen to power. They wouldn't tolerate his/the **genocide** of so many minority populations.
19. The doormen let the **student/gentleman** into the apartment complex. They talked about his/the **studio** in the basement of the building.
20. The donors watched the **gentleman/student** at the charity drive. They suspected that his/the **generosity** was all for show.
21. The doctors told the **painter/teenager** to prepare for the worst after the operation. They hoped his/the **painkillers** would help ease his suffering.
22. The students glared at the **intern/advisor** who accidentally sat in on the meeting. They found his/the **interruption** really rude and annoying.
23. The neighbors knew the **gardener/playboy** was a little eccentric. When they found his/the **garbage** in the middle of the street, however, they felt he had gone too far.
24. The reporters interviewed a(n) **member/artist** of(from) the local performing arts center. They asked about his/the **memoir** which had been recently published.
25. The managers scolded the **bellboy/rocker** for eating too much. They pointed at his/the **belly** bulging out of his skinny jeans.
26. The servant girls congratulated the **prince/cook** on the amateur artwork. They thought that his/the **paintings** were so beautiful they should be put up in all the rooms of the castle

27. The con men watched the **grandpa/senator** react to the news about the fake sweepstakes. They smiled at his/the **gratitude**, knowing there was really no check in the mail.
28. Most people ignored the **warden/psycho** who was known to make dire predictions. They laughed about his/the **warning** about the massive prisoner escape plot being hatched.
29. Many friends helped the **wrestler/realtor** at the scene of the car accident. They gazed at his/the **wreck** down in the ditch and shook their heads in sympathy.
30. The women avoided the **deputy/singer** who was accused of being a sex offender. They whispered about his/the **depravity** and refused to speak to him.
31. The maintenance men told the **singer/deputy** there would be a delay. They said that his/the **sink** wouldn't be installed until next month.
32. Some wolves had been troubling a **shepherd/resident** in the local village. They had raided his/the **shelter** in the woods and eaten all his food.
33. The environmentalists criticized the **policeman/engineer** who was caught dumping toxic waste. They claimed that his/the **pollution** set a poor example for citizens.
34. The ladies calmed down the **waiter/baron** who had just knocked over a tray full of drinks. They said that his/the **wailing** was disrupting the guests.
35. The prosecutors frowned at the **admiral/surgeon** after the confession. They suspected that his/the **admission** was really just a way of covering someone else's tracks.
36. The trainers evaluated the **swimmer/recruit** after a hard day's work. They said that his/the **swiftness** of movement was the key to success.
37. The secretaries complained about the **commander/director** who sent unintelligible memos. They couldn't understand his/the **comments** even when they were very short.
38. The players brought the **coach/driver** back the missing belongings. They had found his/the **coat** on one of the bus seats along with a pair of sunglasses.
39. The royal family praised the **emperor/citizen** for the well-delivered speech. They commended his/the **emphasis** on every word and his clear intonation.
40. The clerks saw the **executive/salesman** hurry into the meeting late. They snickered at his/the **exertion** after running all the way to the office.
41. The police questioned the **professor/boyfriend** of the recently murdered girl. They were shocked by his/the **profanity** when they finally convinced his to talk.
42. The reporters kicked the criminal/genius out of the office after the interview. They couldn't tolerate his/the **criticism** of women and blatant sexism.
43. The colleagues asked the **boxer/critic** to tidy up the shared workspace. They complained that his/the **boxes** were cluttering the space.
44. The party guests avoided the **acrobat/sheikh** from the foreign country. They could smell his/the **acrid** from across the room.

45. The family visited the **actor/pilot** after the heart attack. They insisted that his/the **actuary** draw up a new life insurance policy.
46. The diplomats asked the **ambassador/emperor** to speak more plainly. They said that his/the **ambiguous** demands needed to be clarified.
47. The neighbors reported the **architect/amateur** for threatening public safety. They claimed that his/the **archery** practice was not safe and should be moved elsewhere.
48. The committee praised the **astronaut/cripple** who took the voyage to Mars. They said that his/the **astonishing** goals were an inspiration to all.
49. No one talked to the **barista/skydiver** for very long. They were annoyed by his/the **barrage** of constant questions.
50. Several friends visited the **bartender/counselor** last weekend. They really enjoyed his/the **barbecue** and were impressed with his cooking skills.
51. Few passersby help the **beggar/watchman** who sleeps in the street. They are put off by his/the **begrudging** thanks for their efforts to help.
52. The girls didn't like the **bully/stud** who was on lifeguard duty. They took away his/the **bullhorn** so that he couldn't yell at them any more.
53. Some bears attacked the **camper/twit** in the woods. They left behind his/the **camcorder**, which documented his grisly death.
54. The nurses pitied the **candidate/lunatic** in the hospital. They knew that his/the **cancer** was eating away at his brain.
55. The executives scolded the **workman/slacker** about the company's finances. They fired up his/the **worries** about imminent downsizing.
56. The doctors tended to the conductor/philosopher in the hospital. They promised that his/the contusion would eventually heal.
57. The villagers mocked the **cowboy/reverend** who was scared of thunder. They laughed at his/the **cowardice** and couldn't take him seriously.
58. Several customers sued the **farmer/psychic** for selling bogus merchandise. They claimed his/the **farce** had gone on quite long enough.
59. The judges criticized the **florist/performer** in the competition. They said his/the **florid** shirt simply distracted from his display.
60. The audience booed the **impostor/beginner** who was doing impressions. They disliked his/the **impudence** and tried to throw him off the stage.
61. Most investors liked the **hunk/heir** with expertise in stocks. They relied on his/the **hunches** to decide where to invest.
62. The fans booed the **quarterback/substitute** during the game. They thought his/the **quarrel** with the ref was going to get him kicked off the field.
63. The villagers ignored the **warlock/analyst** before the storm. They believed that his/the **warning** about the impending flood was bogus.

64. The teachers punished the **surfer/wretch** for bad behavior. They wouldn't tolerate his/the **surly** attitude in the classroom any longer.

Semantic conditions

1. Several publishers approached the **journalist/congressman** who had recently made headlines. They said that his/the **article** which had been recently published had stirred up public opinion.
2. The maids lectured the **carpenter/journalist** for being terribly messy. They stared at his/the **overalls**, which were stained and caked with mud.
3. The military prosecutors questioned the **therapist/civilian** during the trial. They focused on his/the **notes** which documented a soldier's misconduct.
4. Many people envied the **photographer/therapist** from the first floor. They that his/the **studio** was very large and had lots of windows.
5. Some housewives visited the **minister/hunter**, who had missed Bible studies. They commented that his/the **congregation** had been concerned about his failing health.
6. The townspeople mourned the **reverend/cowboy** who had died recently. They laid out his/the **Bible** in the coffin with him, as he had requested.
7. Several colleagues criticized the **manager/player** for the failed ad campaign. They worried that his/the **company** was losing value in the stock market.
8. The receptionists chatted with the **attorney/manager** in the office. They asked about his/the **lawsuit** against the music industry.
9. The CIA agents scrutinized the **chairman/terrorist** for the suspicious transactions. They feared that his/the **committee** was merely a front for a more nefarious plot.
10. The TV producers insulted the **reporter/champion** who had become famous during the Olympics. They knew that his/the **interview** would be long and tedious.
11. The actresses criticized the **chauffeur/bodyguard** for having a bad attitude. They got into his/the **car** and complained the entire ride home.
12. The agents arrested the **sheriff/killer** after searching the house. They claimed that his/the **badge** on the desk connected him to the victim.
13. The hostage negotiators freed the **captain/doctor** with clever trickery. They switched out his/the **uniform** from the prison with a lab coat.
14. Some robbers attacked a **knight/priest** who was traveling to Avignon. They struck his/the **sword** out of the palm of his hand and demanded all his money.
15. The directors relocated the **officer/agent** after the war. They thought that his/the **station** in Iraq would no longer be necessary.

16. The plaintiffs argued with the **judge/lawyer** after the trial. They thought that his/the **verdict** couldn't have been arrived at fairly.
17. Many critics admired the **producer/graduate** from the acclaimed art school. They referred to his/the **film**, which had won many awards.
18. The rebels deposed the **general/president** who had recently risen to power. They threw back his/the **army** in a series of decisive and bloody battles.
19. The doormen let the **student/gentleman** into the apartment complex. They searched through his/the **backpack** to make sure he wasn't carrying anything dangerous.
20. The donors watched the **gentleman/student** at the ticket booth. They suspected that his/the **monocle** was all for show.
21. The doctors told the **painter/teenager** to prepare for the worst after the operation. They hoped his/the **pictures** from the trip he had taken with friends would cheer him up.
22. The students disliked the **intern/advisor** who took notes in every meeting. They complained about his/the **typing** because the keyboard on his laptop was very loud.
23. The neighbors knew the **gardener/playboy** was a little eccentric. When they found his/the **flowers** in the middle of the street, however, they felt he had gone too far.
24. The girls met a(n) **member/artist** of(from) the local performing arts center. They asked whether his/the **dues** were terribly exorbitant.
25. The managers scolded the **bellboy/rocker** for eating too much. They pointed to his/the **foodcart**, which he had nearly emptied.
26. The visitors congratulated the **cook/prince** on the renovations recently made to the palace. That thought that his/the **kitchen** had turned out particularly lovely.
27. The women helped the **grandpa/senator** get ready for the reunion. They picked up his/the **bifocals** from the floor, where he had dropped them last night.
28. The security agents searched the **warden/psycho** at the political rally. They took away his/the **baton** so that he couldn't hurt anyone with it.
29. Many friends helped the **wrestler/realtor** at the scene of the car accident. They asked about his/the **trainer**, who had also been in the car.
30. The detectives asked the **deputy/singer** permission to search the house. They found his/the **gun** beneath the bed which he did not have a license for.
31. The maintenance men told the **singer/deputy** about a problem. They had broken his/the **piano** and would have to repair that first.
32. Some wolves had been troubling a **shepherd/resident** at the farm. They had attacked his/the **goats** and even killed one just last week.
33. Many coworkers told the **policeman/engineer** to eat more healthily. They exclaimed that his/the **doughnuts** alone were expanding his waistline by an inch a year.

34. The ladies calmed down the **waiter/baron** who had tripped into the dining room. They picked up his/the **tray** full of dropped wine glasses and asked him if he was okay.
35. The military tribunal punished the **admiral/surgeon** after the war. They stripped his/the **post** away from him and sent him back to civilian life.
36. The trainers evaluated the **swimmer/recruit** after a hard day's work. They said that his/the **goggles** would need to be replaced.
37. The secretaries complained about the **commander/director** who sent unintelligible memos. They couldn't understand his/the **military** jargon and often disregarded his orders entirely.
38. The players brought the **coach/driver** back the missing belongings. They had found his/the **whistle** on one of the bus seats along with a pair of sunglasses.
39. The royal family praised the **emperor/citizen** responsible for the recent policy changes. They said that his/the **kingdom** had really prospered as a direct result.
40. The clerks saw the **executive/salesman** hurry into the meeting late. They stared at his/the **Porsche** longingly after he parked it right out front.
41. The police questioned the **professor/boyfriend** of the recently murdered girl. They were shocked by his/the **eloquence** when they finally convinced him to talk.
42. The reporters kicked the **criminal/genius** out of the office after the interview. They refused to listen to his/the **scheme** to take over the media.
43. The colleagues asked the **boxer/critic** to tidy up the shared workspace. They complained that his/the **gloves** stank and needed to be taken home.
44. The party guests avoided the **acrobat/sheikh** from the foreign country. They sneered at his/the **tumbling** across the room and found him altogether ridiculous.
45. The assistants helped the **actor/pilot** get ready. They worried that his/the **costume** would no longer fit.
46. The diplomats asked the **ambassador/emperor** to speak more plainly. They said that his/the **negotiations** couldn't proceed without clarification.
47. The city council chastised the **architect/amateur** for threatening public safety. They worried that his/the **construction** project wasn't adhering to proper building code.
48. The committee praised the **cripple/astronaut** who was hurt on the voyage to Mars. They said that his/the **disability** only made the achievement more remarkable.
49. No one talked to the **barista/skydiver** for very long. They couldn't stand his/the **coffee** breath, and so avoided talking to him.
50. Several friends congratulated the **bartender/counselor** at the dinner party. They asked for his/the **martini** recipe afterwards because they were so impressed.
51. Few passersby still help the **beggar/watchman** who sleeps in the street. They are put off by his/the **pleading** and mostly ignore him.

52. The girls didn't like the **bully/stud** who was on lifeguard duty. They began ignoring his/the **threats** to kick them out of the pool and eventually he left them alone.
53. Some bears attacked the **camper/twit** in the woods. They tore through his/the **tent** to get at the canned food he had brought.
54. The nurses pitied the **lunatic/candidate** in the hospital. They knew that his/the **insanity** was the result of a degenerative brain disease.
55. The executives called the **slackster/workman** in for a conference. They warned that his/the **procrastination** would no longer be tolerated.
56. The doctors tended to the **conductor/philosopher** in the hospital. They were told that his/the **train** had derailed, nearly crushing him to death.
57. The townspeople cheered the **cowboy/reverend** who won the race. They said his/the **horseback** riding was absolutely superb.
58. Many neighbors doubted the **psychic/farmer** during the drought. They feared that his/the **prediction** for a good rainfall would turn out to be total rubbish.
59. The judges criticized the **florist/performer** in the competition. They said his/the **bouquets**, given to him by friendly admirers, were hardly well deserved.
60. The audience booed the **impostor/beginner** who was doing impressions. They didn't care for his/the **fakery** and tried to throw him off the stage.
61. The bachelorettes competed over the **hunk/heir** at the party. They couldn't resist his/the **handsome** face and flirted with him all night.
62. The fans booed the **quarterback/substitute** during the game. They thought that his/the **fumble** was likely to cost them the match.
63. The villagers ignored the **warlock/analyst** before the storm. They didn't trust his/the **arcane** knowledge even when he was trying to help them.
64. The teachers punished the **surfer/wretch** for bad behavior. They indicated that his/the **flipflops** were in violation of school dress code.

Bibliography

- Altmann, G.T.M., van Nice, K., Garnham, A., & Henstra, J.A. (1998). Late closure in context. *Journal of Memory and Language*, 38, 4, 459–484.
- Anderson, A., Garrod, S. C. & Sanford, A. J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 35, 3, 427–440.
- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22, 261–295.
- Anderson, J. R. (2005) Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science*, 29,3, 313–341.
- Arnold, J., Eisenband, J., Brown-Schmidt, S., & Trueswell, J. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, 76, B13–B26.
- Asher, N. (1993). *Reference to Abstract Entities*. Dordrecht: Kluwer Academic Publishers.
- Baayen, H. R., Piepenbrock, R., & Van Rijn, H. (1993). The CELEX lexical database. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania. Available from WebCelex at <http://celex.mpi.nl/>.
- Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Badecker, W., Miozzo, M. & Zanuttini, R. (1995) The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition*, 57, 193–216.
- Balota, D.A., Yap, M.J., Cortese, M.I., Watson, J.M. (2008). Beyond response latency: An RT distributional analysis of semantic priming. *Journal of Memory & Language*, 59, 495–523.
- Barber, H., & Carreiras, M., (2005) Grammatical Gender and Number Agreement in Spanish: An ERP Comparison. *Journal of Cognitive Neuroscience*, 17, 1, 137–153.

- Barr D.J., Levy R., Scheepers C., & Tily, H. (2013) Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 3, 255–278.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-5. <http://CRAN.R-project.org/package=lme4>.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. <http://CRAN.R-project.org/package=lme4>.
- Bever, T.G. (1970). The cognitive basis for linguistic structures. In J.R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–360). New York: Wiley.
- Bijeljac-Babic, R., Biardeau, A., & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory & Cognition*, 25, 447–457.
- Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23, 45–93.
- Bock, J. K., Carreiras, M., & Meseguer, E. (2012). Number meaning and number grammar in English and Spanish. *Journal of Memory and Language*, 66, 17–37.
- Cacciari, C., Carreiras, M., & Barbolini-Cionini, C. (1997). When words have two genders: Anaphor resolution for Italian functionally ambiguous words. *Journal of Memory and Language*, 37, 517–532.
- 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 & Language*, 56, 103–128.
- Caplan, D., & Waters, G. (2013). Memory mechanisms supporting syntactic comprehension. *Psychological Bulletin and Review*, 20, 2, 243–268.
- Caplan, D., Waters, G. S., & Howard, D. (2012). Slave Systems in Verbal Working Memory. *Aphasiology: Special Issue on Neuropsychology of Short Term Memory*, 26, 279–316.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14, 177–208.
- Caramazza, A. & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: evidence from the ‘tip-of-the-tongue’ phenomenon. *Cognition*, 64, 309–343.

- Carreiras, M. (1992). Parsing strategies in sentence processing: Early versus late closure. *Cognitiva*, 4, 3–27.
- Carreiras, M., & Clifton, C. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech*, 36, 353–372.
- Carreiras, M., Garnham, A., & Oakhill, J. (1993). The use of superficial and meaning-based representations in interpreting pronouns. Evidence from Spanish. *European Journal of Cognitive Psychology*, 5, 1, 93–116.
- Carroll, P., & Slowiaczek, M. L. (1986). Constraints on semantic priming in reading: A fixation time analysis. *Memory & Cognition*, 14, 509–522.
- Chomsky, N. (1981). Lectures on government and binding. Dordrecht, The Netherlands: Foris.
- Clifton, C., Jr., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension. *Rivista di Linguistica*, 11, 11-39.
- Cloitre, M. & Bever, T.G. (1988). Linguistic anaphors, levels of representation, and discourse. *Language & Cognitive Processes*, 3, 293–322.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic.
- Collins, A.M., & Loftus, E.F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–428.
- Colombo, L. (1986). Activation and inhibition with orthographically similar words. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 226–234.
- Corbett, A. T., & Dosher, B. A. (1978). Instrument inferences in sentence encoding. *Journal of Verbal Learning and Verbal Behaviour*, 17, 479–491.
- Cotter, C. A. (1984). Inferring direct objects in sentences: Some implications for the semantics of verbs. *Language and Speech*, 27, 25–45.
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185.
- Cowan, N. (2005). *Working memory capacity*. Psychology Press, Hove, East Sussex: UK.
- Crocker, M.W. (1992). A logical model of competence and performance in the human sentence processor. (Research Paper HCRCJRP-34). Glasgow, UK: Human Communication Research Centre.

- Cuetos, F., & Mitchell, D. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, 30, 73–105.
- Davidson, D. (1967). The logical form of action sentences. In Nicolas Rescher (Ed.), *The logic of decision and action*, 81–95. Pittsburgh: University of Pittsburgh Press.
- Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 668–687.
- Davis, C. J., & Taft, M. (2005). More words in the neighborhood: Interference in lexical decision due to deletion neighbors. *Psychonomic Bulletin & Review*, 12, 904–910.
- Davis, C. J., Perea, M., & Acha, J. (2009). Re(de)fining the orthographic neighborhood: The role of addition and deletion neighbors in lexical decision and reading. *Journal of Experimental Psychology: Human Perception & Performance*, 35, 1550–1570.
- De Moor, W., & Brysbeart, M. (2000). Neighborhood-frequency effects when primes and targets are of different lengths. *Psychological Research/Psychologische Forschung*, 63, 159–162.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M. & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104, 801–38.
- DeLong, K., Urbach, T.P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 8, 1117–1121.
- Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (in press). Contrasting interference profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*.
- Dutke, S. (2003). Anaphor resolution as a function of spatial distance and priming: exploring the spatial distance effect in situation models. *Experimental Psychology*, 50, 4, 270–284.
- Eberhard, K.M., Cutting, J.C., & Bock, J.C. (2005) Making syntax of sense: Number agreement in sentence production. *Psychological Review*, 112, 531–559.
- Ehrlich, K. (1983). Eye movements in pronoun assignment: A study of sentence integration. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes*. New York: Academic Press.

- Ehrlich, K., & Johnson-Laird, P.N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning & Verbal Behavior*, 21, 296–306.
- Federmeier, K.D., & Kutas, M., (1999). A rose by any other name: long-term memory structure and sentence processing. *Journal of Memory and Language*, 41, 469–495.
- Finocchiaro, C., & Caramazza, A. (2006). The production of pronominal clitics: Implications for theories of lexical access. *Language and Cognitive Processes*, 21(1-3), Special issue: Language production across the life span, 141–180.
- Fitzgerald, S. (1925). *The Great Gatsby*. New York: Charles Scribner's Sons.
- Forster, K.I. (1976). “Accessing the mental lexicon”. In R.J. Wales & E. Walker (eds), *New Approaches to Language Mechanisms*, Amsterdam: North Holland.
- Franck, J., Vigliocco, G., & Nicol, J. (2002). Attraction in sentence production: The role of syntactic structure. *Language and Cognitive Processes*, 17, 4, 371–404.
- Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U.H. (2008). The interplay of syntax and form in language production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes*, 23, 3, 329–374.
- Frazier, L. (1985). Syntactic complexity. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language processing: psychological, computational, and theoretical perspectives* (pp. 129–189). Cambridge, UK: Cambridge University Press.
- Frazier, L. (1987). Sentence processing. In M. Coltheart (Ed.), *Attention and performance XII: The Psychology of reading* (pp. 559–586). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Frazier, L., Henstra, J. & Flores d' Arcais, G.B. (1996). Finding candidate antecedents: Phrase or Conceptual Entities. *University of Massachusetts Occasional Papers in Linguistics*, 19, 193–238.
- Garnham, A. (2001). *Mental models and the interpretation of anaphora*. East Sussex: Psychology Press.
- Garnham, A., Oakhill, J., Erlich, M. F., & Carreiras, M. (1995). Representations and processes in the interpretation of pronouns: New evidence from Spanish and French. *Journal of Memory and Language*, 34, 41–62.

- Garrett, M. F (1980) Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production: Vol. I. Speech and talk*, Academic Press.
- Garrod, S. & Sanford, A. J. (1982). The mental representation of discourse in a focused memory system: Implications for the interpretation of anaphoric noun-phrases. *Journal of Semantics*, 1, 21–41.
- Garrod, S. & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language*, 42, 526–544.
- Garrod, S., Freudenthal, D. & Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in discourse. *Journal of Memory and Language*, 33, 39–68.
- Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, 32, 99–156.
- Gibson, E., & Schütze, C.T. (1999). Disambiguation preferences in noun phrase conjunction do not mirror corpus frequency. *Journal of Memory and Language*, 40, 263–279.
- Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: the perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14, 225–248.
- Gibson, E., Pearlmuter, N., Canseco-Gonzalez, E. and Hickock, G. (1996). Recency preferences in the human sentence processing mechanism. *Cognition*, 59, 23–59.
- Gilboy, E., Sopena, J.M., Clifton, C., Jr., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English compound NPs. *Cognition*, 54, 131–167.
- Glaser, W. R. & Düngelhoff, F.J. (1984) The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 640–54.
- Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention. *Cognitive Science*, 17, 311–347.
- Gordon, P.C., Hendrick, R., & Johnson, M. (2001). Memory interference during sentence processing. *Psychological Science*, 13, 425–430.
- Gordon, P.C., Hendrick, R., & Levine, W. (2002). Memory-load interference in syntactic processing. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 27, 1411–1423.

- Grainger, J., & Ferrand, L. (1994). Phonology and orthography in visual word recognition: Effects of masked homophone primes. *Journal of Memory and Language*, 33, 218–233.
- Hartsuiker, R. J., Schriefers, H. J., Bock, J. K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject–verb agreement. *Memory & Cognition*, 31, 1316–1326.
- Hemforth, B., Konieczny, L., Scheepers, C., & Strube, G. (1998). Syntactic ambiguity resolution in German. In D. Hillert (Ed.), *Sentence Processing: A Crosslinguistic Perspective. Syntax and Semantics*, vol. 31 (pp. 292–312). San Diego, CA: Academic Press.
- Hofmeister, P. (2010). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, 26, 3, 376–405.
- Inoue, A., & Fodor, J.D. (1995). Information-paced parsing of Japanese. In R. Mazuka & N. Nagai (Eds.), *Japanese sentence processing*, (pp. 9–63). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Jarvella, R. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, 10, 409–416.
- Jescheniak, J. D. (1993) *Word frequency effects in speech production*. Doctoral dissertation, University of Nijmegen.
- Jescheniak, J. D. & Levelt, W. J. M. (1994) Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Language, Memory, and Cognition*, 20, 824–43.
- Jescheniak, J. D., Schriefers, H., & Hantsch, A. (2001). Semantic and phonological activation in noun and pronoun production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1058–1078.
- Jia, G., Aaronson, D., & Wu, Y. (2002). Long-term language attainment in bilingual immigrants: Predictive variables and language group differences. *Applied Psycholinguistics*, 23, 599–621.
- Jiang, N. (2004) Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25, 603–634.
- Jiang, N. (2007) Selective Integration of Linguistic Knowledge in Adult Second Language Learning. *Language Learning*, 57, 1, 1–33.

- Jones, D. M., Hughes, R. W., & Macken, W. J. (2007). The phonological store abandoned. *Quarterly Journal of Experimental Psychology*, 60, 505–511.
- Jones, D., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 30, 656–674.
- Jonides, J., Lewis, R.L., Nee, D.E., Lustig, C.A., Berman, M.G., & Moore, K.S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, 59, 193–224.
- Kail, M. (1989). Cue validity, cue cost, and processing types in French and Spanish. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of language processing* (pp. 77–117). New York: Cambridge University Press.
- Kazanina, N., Lau, E.F., Lieberman, M., Phillips, C., & Yoshida, M. (2007). Effect of syntactic constraints on long-distance dependency formation in backwards anaphora processing. *Journal of Memory and Language*, 56, 384–409.
- Kennison, S. M. (2004). The effect of phonemic repetition on syntactic ambiguity resolution: Implications for models of working memory. *Journal of Psycholinguistic Research*, 33, 493–516.
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, 2, 15–47.
- Kirby, K. N., & Gerlanc, D. (2013). BootES: An R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods*, 45, 905–927.
- Kush, D. (2013). *Respecting Relations*. Doctoral dissertation, University of Maryland.
- Kutas, M., DeLong, K.A., & Smith, N.J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In Bar, M. (Ed.) *Predictions in the Brain: Using Our Past to Generate a Future*, Oxford University Press, 190–207.
- Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0–6. <http://CRAN.R-project.org/package=lmerTest>.
- Lau, E. F., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of prediction in rapid syntactic analysis. *Brain and Language*, 98, 74–88.
- Leiman, J.M. (1982). A chronometric analysis of referent assignment to pronouns. Ph.D. dissertation, Wayne State University, Detroit, Michigan.

- Lesgold, A., Roth, S., & Curtis, M. (1979). Foregrounding effects in discourse comprehension. *Journal of Verbal Learning & Verbal Behavior*, 18, 291–308.
- Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, 98, 13464–13513.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. Target paper for *Behavioral and Brain Sciences*, 22, 1–75.
- Levelt, W.J.M. (1997). The Genetic Perspective in Psycholinguistics or Where Do Spoken Words Come From? *Journal of Psycholinguistic Research*, 27, 2, 167–180.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 375–419.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10, 44–54.
- Lorimor, H., Bock, J. K., Zalkind, E., Sheyman, A., & Beard, R. (2008). Agreement and attraction in Russian. *Language and Cognitive Processes*, 23, 769–799.
- Lucas, M.M., Tanenhaus, M.K., & Carlson, G. N. (1990). Levels of representation in the interpretation of anaphoric reference and instrument inference. *Memory and Cognition*, 18, 611–631.
- Lupker, S. J., & Columbo, L. (1994). Inhibitory effects in form priming-evaluating a phonological competition explanation. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 437–451.
- MacDonald, M. C., & MacWhinney, B. (1990). Measuring inhibition and facilitation from pronouns. *Journal of Memory and Language*, 29, 469–492.
- MacDonald, M.C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language*, 32: 692–715.
- MacWhinney, B. (1987). Cues, competition, and learning. In J. Miller (Ed.), *Wisconsin Papers on Language Development*. Madison, Wisconsin: University of Wisconsin.
- MacWhinney, B. (2001). The Competition Model: The input, the context, and the brain. In P. Robinson (Ed.), *Cognition and Second Language Instruction*. New York: Cambridge University Press.
- MacWhinney, B., & Bates, E. (Eds.) (1989). *The cross-linguistic study of sentence processing*. New York: Cambridge University Press.

- MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23, 127–150.
- Mancini, S., Molinaro, N., & Carreiras, M. (2013). Anchoring Agreement. In M. Sanz, M. Tanenhaus, M. & I. Laka (Eds.), *Language Down the Garden Path*. Oxford: Oxford University Press.
- Martin, C.D., Thierry, G., Kuipers, J-R., Boutonnet, B., Foucart, A., & Costa, A., (2013) Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, 69, 574–588.
- Mazuka, R., & Lust, X. (1990). On Parameter Setting and Parsing: Predictions for Cross-Linguistic Differences in Adult and Child processing. In L. Frazier, and J. de Villiers (Eds.), *Language Processing and Language Acquisition* (pp. 163–205). Dordrecht: Kluwer Academic Publishers.
- McDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics*, 21, 395–423.
- McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, 29, 111–123.
- McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 817–35.
- McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation*, Vol. 46. San Diego: Academic Press.
- McElree, B., & Dosher, B. (1989). Serial position and set size in short-term memory: time course of recognition. *Journal of Experimental Psychology: General*, 18, 346–373.
- McElree, B., Foraker, S. & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48, 67–91.
- McKoon, G., & Ratcliff, R. (1981). The comprehension processes and memory structures involved in instrumental inference. *Journal of Verbal Learning and Verbal Behavior*, 20, 671–682.
- McKoon, G., Ratcliff, R., & Ward, G. (1994). Testing theories of language processing: An empirical investigation of the on-line lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1219–1228.

- Meyer, A. S., & Bock, K. (1999). Representations and processes in the production of pronouns: Some perspectives from Dutch. *Journal of Memory and Language*, 41, 2, 281–30.
- Mitchell, D.C., & Cuetos, F. (1991). The origins of parsing strategies. In C. Smith (Ed.), *Current issues in natural language processing* (pp. 1–12). Austin, TX: University of Austin Press.
- Mitchell, D.C., Cuetos, F., & Zagar, D. (1990). Reading in different languages: Is there a universal mechanism for parsing sentences? In D. Balota, G.B. Flores d' Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 285–302). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Morris, R. K., & Folk, J. R. (1998). Focus as a contextual priming mechanism in reading. *Memory & Cognition*, 26, 1313–1322.
- Morton, J. (1979). “Word Recognition”. In J. Morton & J.C. Marshall (Eds.), *Psycholinguistics Series*, vol. 2, *Structures and Processes*. London: Elek.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition*, (pp. 264–336). Hillsdale, N. J.: Erlbaum.
- Nicol, J. (1988). *Coreference processing during sentence comprehension*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge.
- Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject–verb agreement processes in comprehension. *Journal of Memory and Language*, 36, 569–587.
- Nicol, J.L., Fodor, J.D. & Swinney, D.A. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 5, 1229–1238.
- Obata, M., Lewis, R.L., Epstein, S., Bartek, B., & Boland, J. (2010). Featural analysis and short-term memory retrieval in on-line parsing: Evidence for syntactic, but not phonological, similarity-based interference. In *Proceedings of NELS 41: Conference of the North East Linguistics Society*, Philadelphia.
- Parker, D. (2014). *The cognitive basis for encoding and navigating linguistic structure*. Doctoral dissertation, University of Maryland.
- Paterson, K., Alcock, A., & Liversedge, S.P. (2011) Morphological priming during reading: evidence from eye movements. *Language and Cognitive Processes*, 26, 4-6, 600–623.

- Paterson, K.B., Liversedge, S.P., & Davis, C.J. (2009) Inhibitory neighbor priming effects in eye movements during reading. *Psychonomic Bulletin & Review*, 16, 1, 43–50.
- Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41, 427–456.
- Peterson, Philip L. 1982. Anaphoric Reference to Facts, Propositions, and Events. *Linguistics and Philosophy*, 5, 2, 235–276.
- Phillips, C, (2010) Syntax at Age Two: Cross-Linguistic Differences. *Language Acquisition*, 17,1–2, 70–120
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Potter, M.C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language* 29, 633–654.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 86, 446–461.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510–532.
- Rinck, M., & Bower, G. H. (1995). Anaphora resolution and the focus of attention in situation models. *Journal of Memory and Language*, 34, 110–131.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107–42.
- Roelofs, A. (1997). *Syllabification in speech production: Evaluation of weaver*. *Language and Cognitive Processes*, 12, 657–94.
- Sachs, J.S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics* 2, 437–442.
- Sag, I.A. & Hankamer, J. (1976). Deep and Surface Anaphora. *Linguistic Inquiry*, 7, 3, 391–428.
- Sag, I.A. & Hankamer, J. (1984). Toward a Theory of Anaphoric Processing. *Linguistics and Philosophy*, 7, 325–345.

- Schmitt, B.M., Meyer, A.S. & Levelt, W.J.M. (1999). Lexical access in the production of pronouns. *Cognition*, 69, 313–335.
- Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 841–50.
- Segui, J., & Grainger, J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime-target frequency. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 65–76.
- Severens, E., Jansma, B., & Hartsuiker, R. (2008). Morphophonological influences on the comprehension of subject-verb agreement: An ERP study. *Brain Research*, 1228, 135–144.
- Shillcock, R. (1982). The on-line resolution of pronominal anaphora. *Language and Speech*, 25, 385–401.
- Slobin, D.I., (1985). *The crosslinguistic study of language acquisition*. Hillsdale, NJ: Erlbaum.
- Spivey-Knowlton, M., & Tanenhaus, M.K. (1994). Referential context and syntactic ambiguity resolution. In C. Clifton, L. Frazier, and K. Rayner (Eds.), *Perspectives on Sentence Processing* (pp. 415–439). Hillsdale, NJ: Erlbaum.
- Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, 60, 308–327.
- Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition*, 114, 447–454.
- Staub, A. (2011). The effect of lexical predictability on distributions of eye fixation durations. *Psychonomic Bulletin & Review*, 18, 371–376.
- Staub, A., & Clifton, C., Jr. (2006). Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 425–436.
- Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1280–1293.
- Taraban, R., & McClelland, J.L. (1990). Parsing and comprehension. In D. Balota, G.B. Flores D' Arcais, and K., Rayner (Eds.), *Comprehension Processes in Reading* (pp. 231–263). Hillsdale, NJ: Lawrence Erlbaum.

- Trueswell, J.C. and Tanenhaus, M.K. 1994. Toward a constraint-based lexicalist approach to syntactic ambiguity resolution. In C. Clifton, L. Frazier, and K. Rayner (Eds.), *Perspectives on Sentence Processing*. Hillsdale, NJ: Erlbaum.
- Van Berkum, J. J. A. (1997). Syntactic processes in speech production: The retrieval of grammatical gender. *Cognition*, 64, 115–52.
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 3, 443–467.
- Van Dyke, J.A. & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, 65, 247–263.
- van Gompel, R.P.G. & Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition*, 90, 255–264.
- Van Turennout, M., Hagoort, P. & Brown, C. M. (1997) Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 787–806.
- Van Turennout, M., Hagoort, P. & Brown, C. M. (1998) Brain activity during speaking: From syntax to phonology in 40 msec. *Science*, 280, 572–74.
- Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Cross-linguistic evidence from head-final structures. *Language and Cognitive Processes*, 25, 4, 533–567.
- Vasishth, S., Bruessow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, 32, 4, 685–712.
- Vigliocco, G., Antonini, T. & Garrett, M. F. (1997) Grammatical gender is on the tip of Italian tongues. *Psychological Science*, 8, 314–17.
- Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject–verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition*, 61, 261–298.
- Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject–verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, 34, 186–215.
- Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes*, 11, 407–442.

- Vincent, S. B. (1912). The function of the vibrissae in the behavior of the white rat. *Behavioral Monographs*, 1, 5.
- Wagers, M. (2008). *The structure of memory meets memory for structure in linguistic cognition*. Doctoral dissertation, University of Maryland.
- Wagers, M. W. & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology*.
- Wagers, M., Lau, E., & Phillips, C. (2009). Agreement attraction in comprehension: representations and processes. *Journal of Memory and Language*, 61, 206–237.
- Wicha, N.Y., Moreno, E.M., & Kutas, M. (2004). Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, 16, 7, p. 1272–1288.
- Yoshida, M., Dickey, M.W., & Sturt, P. (2012). Predictive Processing of Syntactic Structure: Sluicing and Ellipsis in Real-Time Sentence Processing. *Language and Cognitive Processes*, 28, 3, 272–302.