Predictive language processing in late bilinguals: Evidence from visual-world eye-tracking

by Judith Schlenter

Doctoral dissertation submitted to the Faculty of Human Sciences at the University of Potsdam in fulfillment of the requirements for the degree Doctor of Philosophy in Cognitive Science

Date defended: August 7, 2019

Supervisors:
Claudia Felser, PhD
Sol Lago, PhD

Reviewers:
Claudia Felser, PhD
Theres Grüter, PhD

Published online at the
Institutional Repository of the University of Potsdam:
https://doi.org/10.25932/publishup-43249
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-432498

Abstract

The current thesis examined how second language (L2) speakers of German predict upcoming input during language processing. Early research has shown that the predictive abilities of L2 speakers relative to L1 speakers are limited, resulting in the proposal of the Reduced Ability to Generate Expectations (RAGE) hypothesis. Considering that prediction is assumed to facilitate language processing in L1 speakers and probably plays a role in language learning, the assumption that L1/L2 differences can be explained in terms of different processing mechanisms is a particularly interesting approach. However, results from more recent studies on the predictive processing abilities of L2 speakers have indicated that the claim of the RAGE hypothesis is too broad and that prediction in L2 speakers could be selectively limited. In the current thesis, the RAGE hypothesis was systematically put to the test.

In this thesis, German L1 and highly proficient late L2 learners of German with Russian as L1 were tested on their predictive use of one or more information sources that exist as cues to sentence interpretation in both languages, to test for selective limits. The results showed that, in line with previous findings, L2 speakers can use the lexical-semantics of verbs to predict the upcoming noun. Here the level of prediction was more systematically controlled for than in previous studies by using verbs that restrict the selection of upcoming nouns to the semantic category animate or inanimate. Hence, prediction in L2 processing is possible. At the same time, this experiment showed that the L2 group was slower/less certain than the L1 group. Unlike previous studies, the experiment on case marking demonstrated that L2 speakers can use this morphosyntactic cue for prediction. Here, the use of case marking was tested by manipulating the word order (Dat > Acc vs. Acc > Dat) in double object constructions after a ditransitive verb. Both the L1 and the L2 group showed a difference between the two word order conditions that emerged within the critical time window for an anticipatory effect, indicating their sensitivity towards case. However, the results for the post-critical time window pointed to a higher uncertainty in the L2 group, who needed more time to integrate incoming information and were more affected by the word order variation than the L1 group, indicating that they relied more on surface-level information. A different cue weighting was also found in the experiment testing whether participants predict upcoming reference based on implicit causality information. Here, an additional child L1 group was tested, who had a lower memory capacity than the adult L2 group, as confirmed by a digit span task conducted with both learner groups. Whereas the children were only slightly delayed compared to the adult L1 group and showed the same effect of condition, the L2 speakers showed an over-reliance on surface-level information (first-mention/subjecthood). Hence, the pattern observed resulted more likely from L1/L2 differences than from resource deficits.

The reviewed studies and the experiments conducted show that L2 prediction is affected by a range of factors. While some of the factors can be attributed to more individual differences (e.g., language similarity, slower processing) and can be interpreted by L2 processing accounts assuming that L1 and L2 processing are basically the same, certain limits are better explained by accounts that assume more substantial L1/L2 differences. Crucially, the experimental results demonstrate that the RAGE hypothesis should be refined: Although prediction as a fast-operating mechanism is likely to be affected in L2 speakers, there is no indication that prediction is the dominant source of L1/L2 differences. The results rather demonstrate that L2 speakers show a different weighting of cues and rely more on semantic and surface-level information to predict as well as to integrate incoming information.

Keywords: prediction, L2 sentence processing, RAGE hypothesis, German, visual-world eye-tracking

Contents

Acknowledgments

I would like to acknowledge the contributions of the following people and institutions: First of all, I thank Miriam Ellert, Thomas Weskott, Katja Suckow and the whole PsychLab Team at the Georg-August-Universität Göttingen, where I discovered my interest in psycholinguistics. I thank Theo Marinis, who, in 2014, welcomed me as an intern at the University of Reading and showed me how to conduct experiments with children. The Potsdam Graduate School financed several conference trips. These were great experiences and I hugely benefited from the feedback I received on these occasions. I thank the Commission for Research and Young Researchers of the Faculty of Human Sciences, who supported me with money for the participants. This work was also supported by an Alexander-von-Humboldt Professorship awarded to Harald Clahsen. Finally, I would like to thank the geobra Brandstätter GmbH & Co.KG, who kindly provided me with the Playmobil® figures for Experiment 4.

Further, the work presented in this thesis would have been impossible without

. . . my supervisor Claudia Felser, who saw an interesting idea in my proposal, contributed her own ideas and was the driving force behind everything. Thanks a lot for leaving your office door open, so as your PhD student I could bother you with questions on a regular basis.

. . . my co-supervisor Sol Lago, for her constructive criticism, which kept me from making a lot of mistakes, and for her invaluable advice. "You know, sometimes an idea is like an ugly child, as a parent we start to love it but maybe we should abandon it" – Sol's advice when I desperately tried to develop an experimental design on cataphora, testing prediction and not integration. As cruel as it sounds, it ended a phase of constant frustration and allowed me to start with fresh ideas.

. . . Harald Clahsen, who has trusted me with managing his institute since 2017, and who contributed to this thesis with his comments in our regular lab meetings.

. . . all the fantastic people who are or were at the Potsdam Research Institute for Multilingualism: Anna J., Anna S., Cecilia, Clare, Daria, Gunnar, Ilaria, Jana, Janna, João, Kirill, Laura, Lisa, Michaela, Michela, Mona, Seckin, Serkan, Silke, Sina, Sol, Thea, Vera, Yael and Yulia. I have greatly appreciated your feedback and encouragement. I owe special thanks to Cecilia for the pictures she drew for my experiments and to João for helping me with the statistics and always asking how life is. I gratefully acknowledge the technical support provided by Steffen Klemer, who helped me to display all areas of interest in one plot, and all who have helped me in one way or the other but I have forgotten to mention.

... my flatmate Anja (who was also the inspiration for an experimental item that includes a washing machine), sport partners, family and friends, especially Simone. I am particularly grateful to my oldest sister Claudia, who, as a future teacher, introduced me to the fascinating world of science. Whoever finds it creepy that I drag people into a darkened room with artificial light to track their eye-movements during language processing to find out more about native and non-native processing, you can (to some extent) blame her.

... the people who accompanied this thesis in another way: my niece Lea, who never forgot to smile despite all the pain she must have had, and my friend Max, both of whom passed away in 2012. Thank you for the wisdom you shared with me, which no university or textbook could ever have taught me.

Last but not least, I would like to thank the more than 200 people who took part in my experiments, particularly the children and the Russian native speakers, who also sent their friends and family members. Both groups took part in the eye-tracking experiments, the probably most exhausting of all. You were awesome! After more than four years of experiment designing, testing (around 69 hours in rooms without sunlight), analyzing and writing, while it was an invaluable learning process, it would be a lie not to say I am glad the PhD part is done and I can now pursue further goals within academia and beyond.

"Why," said the Dodo, "the best way to explain it is to do it."

(from Lewis Carrol, Alice's Adventures in Wonderland)

List of Tables

List of Figures

List of abbreviations

Acc	Accusative
AIC	Akaike information criterion
AoA	Age of acquisition
AOI	Area of interest
Dat	Dative
CEFR	Common European Framework of Reference for languages
DO	Direct object
IC	Implicit causality
IH	Interface Hypothesis
EEG	electroencephalography
Elog	Empirical logit
ERP	Event-related potential
fMRI	Functional magnetic resonance imaging
HMRGF	Hierarchical Multi-Representational Generative Framework
Hz	Hertz
L1	First or native language
L2	Second or non-native language
ms	Millisecond
Nom	Nominative
NP	Noun phrase

N400	ERP response; negativity that has its peak approx. 400 ms after stimulus onset
OSV	Object-subject-verb
ot1	First-order (linear) orthogonal time polynomial
ot2	Second-order (quadratic) orthogonal time polynomial
ot3	Third-order (cubic) orthogonal time polynomial
OVS	Object-verb-subject
PO	Prepositional object
PP	Prepositional phrase
P600	ERP response; positivity that has its peak approx. 600 ms after stimulus onset
RAGE	Reduced Ability to Generate Expectations
SD	Standard deviation
SE	Standard error
SMI	SensoMotoric Instruments
SSH	Shallow Structure Hypothesis
SOV	Subject-object-verb
SVO	Subject-verb-object
VW	Visual-world eye-tracking

Chapter 1

Introduction

Most of us know such situations where we can foresee what will happen: We see a clumsy little toddler taking his first steps and we expect him to fall. We have a conversation with someone, and we know how the other person will complete his utterance. This is not because we are magically gifted somehow, but because we have had these experiences many times and we have learned to predict the outcome based on certain cues. Several researchers claim that generating predictions is an essential function of the human brain. Bar (2007) calls it the "proactive brain," while Clark (2013) speaks of the human brain as a "prediction machine." But what is the prerequisite for predicting upcoming information? What sources of information can we use as a predictive cue and what can we predict? For language processing, the sources used to predict are well investigated for native speakers, ranging from the use of prior context and lexical-semantic information to morphosyntactic marking on verbs and NPs or discourse-level cues. However, less is known about prediction in learner populations. Can a person who has learned a further language later in life use the same sources of information as a native speaker of that language? Or what about the little toddler who is acquiring his mother tongue(s)?

Before turning to these questions, I would like to define what will be meant by prediction in language processing throughout the text, as researchers have used this term in different contexts and with different meanings. Here I will use the term as defined by Huettig (2015), who sees predictive processing as the "pre-activation/retrieval of linguistic input before it is encountered by the language comprehender" (p. 122). To be more concrete, I will explain this with two examples, each using a different method: In a famous experiment by DeLong, Urbach, and Kutas (2005), participants read a sentence context like (1) word-by-word, while event-related potentials (ERPs) were recorded from their scalps. When encountering the article *an* in (1-b), indicating that the more likely word *kite* cannot follow, participants showed a different neural response than for (1-a). Crucially, the effect had its onset before the target noun, demonstrating that the participants not only predicted the upcoming noun but also its phonological form.

(1) The day was breezy so the boy went outside to fly ...

  a. a kite

  b. <u>an</u> airplane

A method that manages to measure prediction without any violations of expectation is the visual-world eye-tracking paradigm I will describe in more detail in section 2.1. It offers many advantages compared to other methods and was employed for the experiments in this thesis. A famous visual-world eye-tracking example and one of the first experiments demonstrating prediction was conducted by Altmann and Kamide (1999). It showed that, when listening to the sentence in (2), participants oriented their gaze towards the only edible object in the visual scene, the cake, even before it was mentioned. If participants heard the sentence with a neutral verb like *move* instead, they were less likely to look at the cake. The experiment demonstrated that participants made use of the lexical-semantics of verbs like *eat*.

(2) The boy will eat the cake.

Many experiments that take their findings as evidence for prediction have used a design where an effect was measured only after the presentation of a target word, hence their results can also be viewed as resulting from information integration. Since it is difficult to tease apart rapid integration and prediction, I will take as a primary criterion for prediction the time course of an effect: Only effects visible prior to the onset of the critical perceptual input are taken as effects of prediction. Throughout, I will not discriminate between the different mechanisms underlying prediction, as described in section 2.2 together with other key technical terms, because it seems likely that mechanisms like proactive prediction and prediction through priming, for example through lexical association, interact. This leads me to the second point regarding prediction, namely how it is accomplished and what its purpose is.

A recent framework for prediction is the Hierarchical Multi-Representational Generative Framework of Language Comprehension (HMRGF) by Kuperberg and Jaeger (2016), which is followed throughout the text. It assumes that multiple types of information can be used to probabilistically predict upcoming information at multiple levels of representation. Here, the term *generative* does not follow the Chomskyan tradition but refers to the likelihood of how a sentence will be generated and not whether a sentence could be generated. The comprehender's goal plays an important role in this framework, which typically is to infer the intended message-level interpretation or situation model. This goal can be achieved more rapidly and efficiently if the comprehender incrementally updates the hypothesis about a message on the basis of each new piece of incoming information. If the bottom-up input is incompatible with a prediction generated through inference at high-level representations, i.e., the comprehender has already pre-updated her hypothesis about a message and pre-activation at the respective lower levels of representation has taken place, additional neural processing can be the consequence, which might reflect a process of adaptation or learning. Whether people predict or not can depend on their respective goal, the task at hand and statistical contingencies between stimuli. To maximize the utility of prediction and rationally allocate resources, the comprehender estimates the reliability of prior beliefs and can use this estimate to modulate the degree of belief updating. In other words, if a person reads a sentence like (1), she will update her situation model with each new incoming word, i.e., bottom-up, so after encountering all the information up to the article of the target noun (breezy day, boy, fly), she is likely to have already pre-updated her belief about the event (boy flies kite) and then uses this high-level information to predict the specific lexical item *kite*. This last process can also be described as top-down processing. As could be shown by DeLong et al. (2005), the prediction not only includes semantic features of upcoming lexical items but also phonological information. Depending on the context, the reliability of cues and other factors, including but not limited to the availability of cognitive resources, upcoming information is more or less likely to be predicted.

As a short recap, language comprehenders have been shown to use multiple sources of information to predict upcoming information at multiple levels of representation, for example semantics but also phonology. However, much of what is known about prediction stems from studies on adult native or first language (L1) speakers. Now, I would like to return to the questions asked at the beginning of this introduction: What about late bilinguals and children? Findings from early experimental studies with second language (L2) speakers have led to the assumption that their predictive processing abilities are reduced as compared to L1 speakers. This assumption has become known as the Reduced Ability to Generate Expectations (RAGE) hypothesis (Grüter, Rohde, & Schafer, 2014). In contrast, children acquiring their L1 have been found to predict from very early on. This is shown in the literature review for L1 speakers in section 4.1, which also includes studies that tested children. This apparent dissociation is particularly interesting, because several researchers suggest that there is a link between prediction and language learning (e.g., Altmann, 2002; Altmann & Mirkovic, 2009; Chang, Dell, & Bock, 2006; Chang, Kidd, & Rowland, 2013). In the framework by Kuperberg and Jaeger (2016), the authors also mention learning as the result of previous false prediction. However, as is often the case, it is a chicken-and-egg problem: Do children learn to predict or do they predict to learn? Along these lines, Rabagliati, Gambi, and Pickering (2016) state that it is difficult to disentangle whether prediction is a cause or a consequence of language learning and that "[...] prediction, conversation, and learning may all interact, such that prediction and learning help to explain, how, as adults, we engage in conversation, while conversation, in concert with prediction, helps to explain how children eventually learn" (p. 103). Against this backdrop, the aim of this thesis is to shed further light on the predictive processing abilities of late bilinguals by also comparing the results with what is known for children or, in the last experiment presented, by making a direct comparison. Why is it that L2 speakers' predictive abilities are reduced and is it really a general deficit or is it a selective one? As will be shown in the literature review for L2 speakers in section 4.2, there is a greater indication of the latter. Whereas prediction based on lexical-semantic or contextual information seems to be unproblematic, prediction based on morphosyntactic information or prediction at the discourse-level, requiring the integration of information sources from different linguistic domains, seem to pose a challenge for L2 speakers.

If L2 speakers' predictive processing abilities were only selectively limited, what could be the reason? In section 3.1, I will discuss what differentiates late L2 learners from child L1 learners and, in section 3.2, I will point out L2 processing accounts that may give some indication of why L2 speakers sometimes have been shown not to predict. According to the Shallow Structure Hypothesis (Clahsen & Felser, 2006a, 2006b), L2 speakers may rely less on grammatical than on non-grammatical information in comparison to L1 speakers. Furthermore, they may differ from L1 speakers in their real-time integration of different cues. According to the Interface Hypothesis (e.g., Sorace, 2011; Sorace & Serratrice, 2009), this should be particularly true when the discourse-level is involved, potentially due to cognitive demands. Note that while these hypotheses provide the motivation to test for selective limits of prediction, none of them make any claims about predictive processing. I will also discuss general factors that may affect language processing in the individual and might hinder L2 but also L1 speakers from predicting, including L2-specific factors which should be considered in experimental studies. One such factor is the cognitive resources available. An L2-specific factor, for example, is the experience with a cue, i.e., whether and how a cue is realized in the L1. Hence, I will discriminate between general limits and selective limits and, in the experiments to follow, attempt to tease these apart. The overarching research question followed throughout will be:

Are L2 speakers' predictive processing abilities selectively limited?

To systematically test for selective limits, I conducted four visual-world eye-tracking experiments in German. These examined the following predictive cues: (1) the lexical-semantics of verbs, (2) verb number marking, (3) case marking and (4) implicit causality. This way, I included a linguistic phenomenon that can be assigned to the lexicon, two phenomena that involve morphosyntactic operations and a phenomenon including the discourse-level. I expected to find an effect of prediction for a lexical-semantic cue for L2 speakers, but their predictive abilities might be limited when it comes to morphosyntactic and discourse-level cues.

Experiment 1: Verb semantics and the selectional restrictions for upcoming arguments are assumed to be stored in the mental lexicon. To control whether L2 speakers use this source of information in the same way as L1 speakers, it was tested whether verb information was used to restrict the choice of an upcoming noun to the semantic category animate or inanimate, which also involves world knowledge, before the noun was encountered. It was expected that L2 speakers have no problem with semantic information and show a nativelike pattern.

Experiment 2: The investigation of subject-verb number agreement is interesting for several reasons. Although it has been shown to be an early acquired cue and is used predictively by children (Lukyanenko & Fisher, 2016), an eye-tracking study by Riordan et al. (2015) in English could not detect an effect of prediction in adult L1 speakers. As a possible explanation for their findings the authors mention the reliability of the cue, with number often not being reliable and the more experienced a language user, the more counterevidence she might have encountered that prevents her from using it. This is in conflict with the assumption that the more reliable a cue, the earlier it is acquired (e.g., MacWhinney, 2001). In German, number agreement occurs within NPs, where number marked on the article and/or adjective is highly reliable regarding the number of the upcoming noun, or across phrases. Since agreeing verbs can either precede or follow their subjects, number agreement within an NP appears to be a stronger cue than number agreement across NPs. So far, to my knowledge, it is unclear whether L2 speakers can use number marking on a verb to predict the number feature of the upcoming noun, but it has been shown within NPs (Hopp, 2012). In the context of the mixed findings for adult L1 English speakers, it was additionally interesting to test L1 German speakers' use of verb number marking.

Experiment 3: The use of case marking to predict an upcoming thematic role involves several processing steps: lexical-semantic access, recognition of case and a successful syntax-semantics mapping. Previous findings have indicated that these operations are difficult for L2 speakers (Hopp, 2015; Mitsugi & MacWhinney, 2016). However, previous studies have tested L2 speakers whose L1 English has no proper case marking system, so it is unclear which role language similarity might play. Here, an L2 group was tested whose L1, like German, has a rich case marking system, namely Russian. The use of case marking was tested by manipulating the word order in double object constructions after a ditransitive verb (Dat > Acc vs. Acc > Dat). If L2 speakers, in line with previous findings, are unable to use case, they should show the same gaze pattern for both linearization patterns.

Experiment 4: Since, to my knowledge, there is no information how or when children use discourse-level cues to predict, the use of implicit causality (IC), a phenomenon well investigated for adult L1 speakers, was tested with a group of German-speaking children and late L2 learners of German and the data compared to that from a group of adult L1 speakers. This group comparison enabled me to control for an effect of cognitive resources, in particular (working) memory, since children were expected to be cognitively less mature than grown-ups. An additional digit span task confirmed that the children as a group had a lower memory capacity than the adult L2 speakers they were compared to. Specifically for a phenomenon at the external discourse interface, cognitive resources may play a role and, thus, may affect children and L2 speakers alike. In the case of IC information, information about argument and discourse structure needs to be integrated in order to predict at the discourse-level. If the results still show a difference between the L1 groups and the L2 group, this would be more evidence for L1/L2 differences.

Complementary offline tasks tested whether the L2 speakers were familiar with the cues investigated in the eye-tracking experiments. Moreover, I tried to control for a range of factors that may impact L2 processing. One factor was L1-L2 similarity. To rule out that a cue was not used predictively, either because the L2 group did not have sufficient experience with it, as it does not exist in their L1, or because it is realized in a different way, Russian L1-German L2 speakers were tested. As will be discussed in the introduction to each experiment, these languages are similar in the way they encode the information crucial to predict. An exception is Experiment 2, where the verb and, thus, the number cue in the Russian translation would be left unrealized in the construction under investigation. However, the phenomenon itself, subject-verb number agreement, does exist in Russian. Below, a full list of factors I tried to control for is given:

- Proficiency: All L2 speakers tested were highly proficient; offline tests moreover examined whether a cue was generally known (subject-verb agreement, case marking, implicit causality).
- Language similarity: All cues tested also exist in the L1 of the L2 speakers being tested and do not fundamentally differ in their realization.
- Lexical knowledge: Vocabulary tests after the online experiment tested whether L2 speakers knew the meaning of crucial words, here knowledge of the restrictive verbs (Experiment 1) and implicit causality verbs (Experiment 4).
- Cognitive resources: In Experiment 4, a group of 'cognitively less mature' children was included and compared to the L2 group.

L2 speakers' proficiency was assessed through the Goethe placement test (Goethe-Institut, 2011), a 30-item multiple choice test. The L2 speakers included in the analyses had a score of 21 or above, which corresponds to upper B2- to C1/C2-level according to the Common European Framework of Reference for languages (CEFR, Council of Europe, 2001). Another selection criterion in addition to L2 proficiency was the age of acquisition (AoA) of German. Only L2 speakers were included who started to learn German at or after the age of seven, here defined as late bilinguals. Hence, most participants learned German in school or at university and, thus, not exclusively in the spoken modality as is typical for L1 acquisition. None of the L2 speakers included in the analyses reported having learned a language other than Russian from birth. The German L1 groups included no early bilinguals as well.

This thesis followed two goals: The first and central one was to shed further light on the predictive processing abilities of late L2 speakers and locate the origin of L1/L2 differences by testing for selective limits in terms of the sources of information used to predict. At the same time, this thesis aimed to contribute to what is known about predictive processing in L1 speakers. To the best of my knowledge, no (published) study on prediction has investigated subject-verb number agreement in German, a morphologically richer language than English, or the use of case marking in German double object constructions. These two experiments were, therefore, explorative and, as laid down in the respective discussion sections as well as the general discussion at the end, can provide the basis for further research. Moreover, the study on implicit causality provides a direct comparison between an adult control and two different learner groups, a child L1 and adult L2 group; although this is preferable, such experiments are seldom conducted in psycholinguistic research.

The thesis is structured as follows: First, I will introduce the method employed to measure predictive processing in the experiments conducted, the visual-world eye-tracking paradigm, and specify the key technical terms used throughout. These also include mechanisms that underlie prediction, here used as an umbrella term for all sorts of predictive mechanisms (chapter 2). In the following chapter, I will give a short overview of the characteristics of L1 and L2 acquisition, before turning to the RAGE hypothesis and other L2 processing accounts relevant in this thesis. In this chapter, I try to work out the potential reasons why prediction might be differently affected in L1 and L2 processing (chapter 3). The last background chapter presents a review of the experimental evidence on prediction available so far, covering prediction in L1 speakers – note that here only a sample of studies could be considered that fit the current definition of prediction – and in L2 speakers with a focus on late learners. The literature review will close with a summary and overview of the results from previous studies (chapter 4). The next four chapters cover the experiments conducted, investigating prediction based on the lexical-semantics of verbs (chapter 5), prediction based on verb number marking (chapter 6), prediction based on case marking (chapter 7) and, finally, prediction based on implicit causality information (chapter 8). In chapter 9, I will discuss the results in light of the individual and selective differences proposed in chapter 3, before I come to the conclusion in chapter 10. The three appendices contain a description of the experimental design, the set-up and the data preparation (Appendix A); they also contain the materials and additional information for the first three experiments, which were conducted together in one study (Appendix B), as well as the materials and additional information for Experiment 4 on implicit causality (Appendix C).

Chapter 2

Methods and key technical terms

2.1 The visual-world eye-tracking paradigm

In the past, two methods have been shown to be particularly useful for measuring predictive processing. One method is the measurement of ERPs. In ERP experiments, participants are typically exposed to sentences that include either an expected or an unexpected continuation. If participants predict upcoming input, they should show a different neural response for an unexpected than for an expected continuation prior to the critical input (e.g., in the case of the prediction of the gender feature of an upcoming noun a different neural response that has its onset on the preceding article). Another useful method is the measurement of eye-movements, which, unlike the ERP method, requires no violation of expectation. In the following, I will focus on the method of eye-tracking during listening, especially the visual-world (hereafter 'VW') paradigm, where people are presented with a visual display, a scene or single pictures, while listening to speech. In eye-tracking during listening experiments, eye-movements to a target picture or object before the onset of the critical perceptual input, so called anticipatory eye-movements, indicate a prediction process. A further advantage in addition to its more naturalistic exposure is that eye-tracking during listening provides a well-suited method for testing sentence comprehension in less literate children and L2 speakers whose native language uses a different script; for an overview of the method, particularly its use in L2 research, see also Berends, Brouwer, and Sprenger (2016).

The use of eye-movement monitoring to study spoken language goes back to Cooper (1974), who discovered that people tend to look at depicted objects when they are mentioned or when there is an association with other words in a text. A widespread use of this methodology, however, first began in the 1990s with a study on syntactic ambiguity resolution conducted by Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) published in *Science*. In this study, participants were asked to act out a task: While wearing a head-mounted eye-tracker, they were, for example, instructed to *Put the apple (that's) on the towel in the box*. At the same time, a visual display that contained either only one referent, one apple on a towel, or two referents, one apple on a towel and one on a napkin, was presented to them. Participants' eye-movements revealed whether they interpreted the second NP *on the towel* as the destination or as modifying information and whether this varied depending on syntactic ambiguity and visual context type. The study could demonstrate that people rapidly start to structure the linguistic input and that this process can be influenced by the visual context. It was followed by a range of similar experiments, also including children (e.g., Trueswell, Sekerina, Hill, & Logrip, 1999) and L2 speakers (Pozzan & Trueswell, 2016). In subsection 3.1.1, this experimental design will be described in more detail, as it provides interesting insights into the developmental changes in language processing.

Altmann and Kamide (1999), as briefly described in the introduction, found that when presented with a sentence like *The boy will eat the cake*, participants looked earlier towards the depicted cake than when the verb was instead *move*. The participants oriented their gaze towards the only edible object in the visual scene, the cake, even before it was mentioned, demonstrating anticipation based on the selectional restriction of the verb; an example of such a visual scene is shown in Figure 2.1. Further research showed that people not only use a single cue like the lexical-semantics of verbs to generate such anticipatory eye-movements, but can combine multiple cues (e.g., Borovsky, Elman, & Fernald, 2012; Kamide, Altmann, & Haywood, 2003). Another form of eye-tracking during listening, which is less complex and often used in developmental research with infants and toddlers, is the preferential looking paradigm (Hirsh-Pasek & Golinkoff, 1996). There, participants see only two pictures and their preference for one picture over the other is assessed. VW experiments often include more than two pictures as further described below.

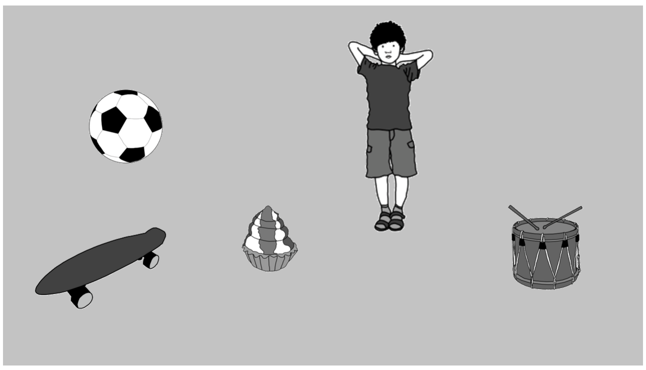


Figure 2.1: Example of a semi-realistic scene used in visual-world eye-tracking, constructed with pictures from the MultiPic database (Dunabeitia et al., 2018)

The observation that the probability of looks to an object increases when it is mentioned or activated, independent of whether the experiment includes an active or passive look-and-listening task, is known as the linking assumption. The method of eye-tracking during listening offers the possibility to measure the underlying processes during language comprehension in a relatively naturalistic way, where the responses can be closely time-locked to the input. A gaze shift is thereby interpreted as a shift in visual attention, which is realized through a saccadic eye-movement. Programming an eye-movement takes around 200 ms (Matin, Shao, & Boff, 1993), which should be considered when analyzing and interpreting eye-movement data. It is unclear whether new information is processed during a saccade. Typical dependent measures are fixation proportions to an area of interest (AOI) or counts of saccades to that area in a predefined time window, which can vary depending on the research question (see, e.g., Boland, 2004; Huettig, Rommers, & Meyer, 2011; Pyykkönen-Klauck & Crocker, 2016; Tanenhaus & Trueswell, 2006). Since here the research question is whether people predict, the critical window covers the segment where anticipatory eye-movements can first be initiated based on the bottom-up input up to the onset of the critical perceptual input (e.g., the target noun). It has to be noted that not only the auditory input guides people's sentence interpretation; a reason why a target is recognized earlier is because the visual presentation already leads to the activation of its phonological representation (Huettig et al., 2011). Important for VW experiments, especially when testing for prediction, is the preview time as shown, for example, by Ferreira, Foucart, and Engelhardt (2013). People need some time to scan the visual display beforehand to be able to predict. A common preview time in VW experiments is around 1000 ms.

Researchers have used different visual displays to investigate eye-movements to depicted objects, ranging from semi-realistic scenes (Figure 2.1) to single pictures that are arranged at a similar distance to each other on the screen like in Borovsky et al. (2012); see Figure 2.2.

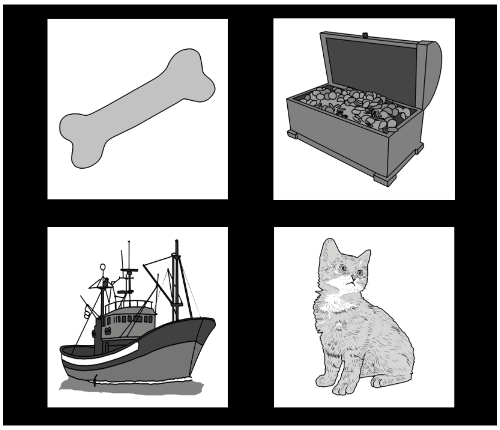


Figure 2.2: Visual display in Borovsky et al. (2012), reconstructed with pictures from the MultiPic database (Dunabeitia et al., 2018)

Here, the researchers were interested in the use of combinatorial information in adults and children. They presented participants with sentences like *The pirate hides the treasure*. The sentence object and target *treasure* thereby refers to one of the depicted objects in the display but the researchers also added an agent-related object, a *ship*, as well as an action-related object, a *bone*. An additional picture of a *cat* is a distractor here but functions as the action-related object in the sentence version *The pirate chases the ship*, the agent-related object in *The dog hides the bone* and the target in the version *The dog chases the cat*. This experiment represents a good example of a balanced design, where the visual display remains the same across experimental conditions and, therefore, potential differences in visual salience are better controlled for. For the experimental findings, see subsection 4.1.4 for adult and child L1 speakers and subsection 4.2.1 for adult L2 speakers. Similarly, Hopp (2013), in an experiment in German, showed participants a four-object display, while three of the depicted objects were matched in color. In 'same' trials, all color-matching objects had the same grammatical gender, while in 'difference' trials one object could be clearly identified based on a gender cue on the article preceding the target noun (e.g., *Wo ist das gelbe Kleid?* – Where is the-[*neut*] yellow dress?). For the analysis, the two trial types were then compared. For the experimental findings, see subsection 4.1.4 for adult and child L1 and subsection 4.2.4 for adult L2 speakers.

These are two examples to illustrate how the visual-world paradigm can be used. Whereas Borovsky et al. (2012) examined whether participants combine information about the Agent and the action and also consider other objects in the display in addition to the target noun, the experiment by Hopp (2013) aimed to reveal differences in reaction times between difference trials, where gender was informative regarding the upcoming noun, and same trials. In this experiment, a fixation cross was used, which the participants were instructed to fixate after a sound signal was heard, to avoid potential baseline effects. In addition, Hopp removed all data in which participants already looked at the target within 200 ms after determiner onset, arguing that such early looks cannot reflect linguistically guided gaze shifts. As an alternative to a fixation cross, researchers have added some kind of gaze neutralizer. For example, Cozijn, Commandeur, Vonk, and Noordman (2011, p. 384) inserted a prepositional phrase (e.g., *in the car*) that was also depicted in the visual display before the critical region started to attract the eye-movements away from the other AOIs. Also depending on the respective research question, there is a debate whether the removal of all on-target trials should be recommended (when interested in reaction times as above it is surely necessary). Barr, Gann, and Pierce (2011) suggest controlling for baseline effects by using statistical estimation together with an experimental design that tries to avoid such effects.

The last aspect already points to a major issue with VW data – there is no consensus on how to analyze them. The problem not only includes the data preparation (removal of data points, etc.) but also the statistical methods. When analyzing VW data, one has to deal with two different variables: the dependent variable looks, i.e., whether a participant looked towards a particular AOI or not, a categorical variable, and the independent variable time, a continuous variable. As pointed out by Barr (2008), traditional analyses like ANOVAs or *t*-tests are designed for the opposite situation, when the effect of a categorical variable on a continuous variable needs to be analyzed like in reaction time experiments. Nevertheless, VW data have been analyzed this way by constructing a series of time bins and calculating a proportion that aggregates over time and multiple trials. The last procedure carries the risk that baseline and rate effects may be ignored. Baseline effects, which Barr calls anticipatory effects, arise when an effect is already present prior to the onset of a critical window. Rate effects are effects that arise as a change in the distribution of looks over time within a critical window. Another issue with VW data is that observations are not independent. The non-independence is a result of the multilevel sampling scheme, a single trial including not only one but many observations, and the mechanics of the eye, it taking time to initiate a gaze shift. Barr (2008) advocates the use of multilevel logistic regression for VW data. The logistic part is required to accommodate the categorical dependent variable, and the regression part to accommodate the continuous predictor variable of time. Instead of using proportions, the dependent variable looks (e.g., fixations) is log-transformed. Whereas the proportion is the ratio of an observation to all observations, the log of the odds is the ratio of a positive occurrence (an event was observed) to negative occurrences (an event was not observed). The log odds or logit scale is unbounded and symmetric around zero whereby zero indicates that a target event was equally likely to occur as not to occur. To filter out the dependencies mentioned above, he suggests aggregating together all trials within an experimental condition, grouping the observations into a series of time bins and then computing the empirical logit for each bin.[1]

One type of multilevel regression that is recently being used more and more is the growth curve analysis. Originally developed for longitudinal data, this analysis enables the researcher to uncover the trajectory of an effect over the time course. To capture change over time, orthogonal time polynomials are included in the model. A description is given by Mirman, Dixon, and Magnuson (2008): Whereas the intercept (0th order) shows the constant difference or the effect for the whole time window defined, the linear term (1st order) indicates a single change in focus, the quadratic term (2nd order) two changes in focus and so on. Note though that interpretation becomes problematic at a certain point. It is useful to include the linear term when we expect that participants will shift their gaze towards a target AOI from a neutral start. If it is expected that participants will first look to a competitor AOI from a neutral start and only then towards the target AOI, the quadratic term should be included. However, which terms need to be included should be hypothesis-driven. Figure 2.3 shows an idealized graph where the onset is at chance level and then looks to target either increase linearly or non-linearly, i.e., there is a curvature. There are several reasons why at the beginning of a critical region looks to an AOI do not start at chance level. This can be, for example, because one picture appears more attractive, as is often the case for animate objects (for a note on visual salience, see also Altmann & Kamide, 1999, p. 254).

The eyetrackingR package (Dink & Ferguson, 2015), which was used here, has implemented a way to calculate such a growth curve analysis. Among other analyses, the package also provides a way to determine the onset of an effect through the calculation of a bootstrapped cluster-based permutation analysis. More information about the data processing for the experiments and the experimental designs is provided in Appendix A.

[1] For empirical logits, a small value is added to the logit equation to account for the case that data points correspond to zero or one. However, the use of empirical logits has not been without criticism as well (see Donnelly & Verkuilen, 2017).

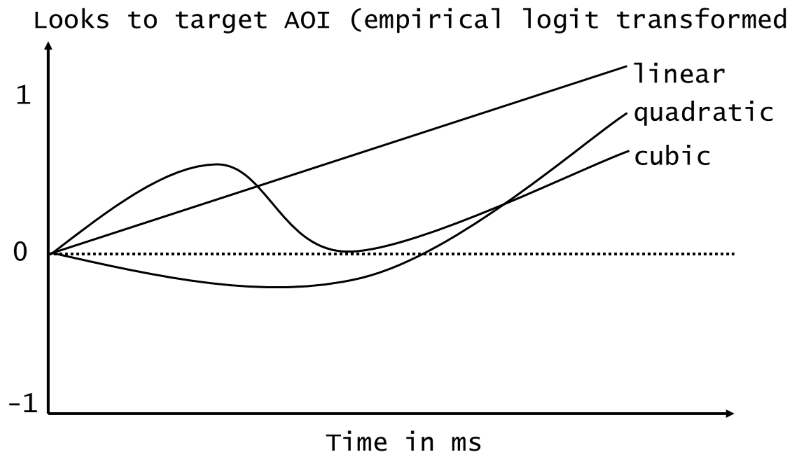


Figure 2.3: Example of a time course graph. The dotted line marks chance level.

2.2 Predictive mechanisms

In this section, I will clarify some technical terms associated with prediction and define the key technical terms used throughout. Many researchers, for instance, use the terms *prediction* and *anticipation* interchangeably; note however that there are subtle differences (e.g., Altmann & Mirkovic, 2009; van Petten & Luka, 2012). Prediction can be described as a processing mechanism that is domain-general and is also used during language processing to predict upcoming linguistic stimuli. However, underlying prediction even more mechanisms are at work, some of which I would like to mention at least briefly, because they will become relevant in the discussion of previous studies or, like the Bayes' rule, are important for the understanding of the concept of prediction. I will further define some terms used in the context of methods employed by researchers testing for effects of prediction in language comprehension.

2.2.1 Prediction, predictability and anticipation

According to DeLong, Troyer, and Kutas (2014) and Kuperberg and Jaeger (2016), the term *prediction* in its traditional sense defines the activation of a linguistic item in advance of perceptual input and is considered an all-or-nothing phenomenon. The process would be conscious, strategic and attention-demanding and could lead to processing advantages but also involves potential costs. Differently, *anticipation* refers to the pre-activation of more than one item and is sometimes used for contextual representations at a more coarse-grained level. For instance, van Petten and Luka (2012) reserve the term *prediction* for the expectation of a specific upcoming word, whereas they use the term *anticipation* for semantic content. They note, however, that a broad semantic expectation may lead to a specific lexical prediction. To illustrate the difference, take as an example an ERP experiment by Federmeier and Kutas (1999). Here, participants read a sentence like (3). The final noun of the last sentence was either the target (3-a), a categorially related (3-b) or an unrelated word (3-c). Throughout, sentences or contexts like these will be called *constrained*, because they constrain the reader/listener towards a specific word or only a set of words.

(3) He caught the pass and scored another touchdown. There was nothing he enjoyed more than a good game of . . .

  a. football (target)
  b. baseball (related)
  c. monopoly (unrelated)

Even though *baseball* and *monopoly* were both implausible continuations, the related word *baseball* induced a smaller N400 than the unrelated word *monopoly*. An N400 is a negative-going waveform peaking around 400 ms after stimulus onset. The amplitude of the N400 component, typically observed after a meaningful stimulus, has been shown to be reduced when the context supports its interpretation (e.g., Kutas, van Petten, & Kluender, 2006). It is argued by van Petten and Luka (2012) that the findings from this related-anomaly experiment would reflect rapid integration rather than the confirmation or disconfirmation of a prediction, i.e., the prediction of the word *football*. Note that it is indeed not a good example of prediction for the reason that the N400 is measured after the onset of the target word. However, many assumptions regarding prediction are based on findings from experiments like these. For example, Kutas and Federmeier (2000) view the results of this experiment as an indication of a predictive strategy: The language comprehender uses all information available as fast as possible to constrain the search in her semantic memory and facilitate the processing of the item(s) most likely to occur. The authors explain the graded effect in terms of the pre-activation of semantic features (e.g., for the example above a game where a ball is involved). On the other hand, van Petten and Luka (2012) attribute the late positive component sometimes observed in ERP experiments to cognitive costs that arise when integration, for example due to failed prediction, is perturbed and re-analysis is required; for a discussion on costs of prediction, see Kutas, DeLong, and Smith (2011). This includes the P600, an ERP component mentioned in the context of the following literature review. Brothers, Swaab, and Traxler (2015) could show that conscious prediction of a specific lexical item (the participants were asked whether the word shown was the one they had predicted) resulted in robust and rapid facilitation during lexical processing, whereas neural responses were slightly different for sentences where the context was only supportive, similar to (3-b) in the example above. Throughout this work, I will not discriminate between the prediction of a specific lexical item and more broader expectations sometimes referred to as anticipation. However, I would like to note at this point that there presumably is a difference between them, at least regarding the time course over which they affect processing.

In line with Kutas and Federmeier (2000), prediction has more recently been seen as a graded and probabilistic phenomenon that presumably proceeds in parallel, so more than one continuation is anticipated at the same time but with a different strength (Kuperberg & Jaeger, 2016). This leads me to the next term, namely that of *predictability* or *expectancy* (e.g., DeLong et al., 2014). This term describes the likelihood that an item is predicted from a given context, which is typically assessed through a cloze test. In a cloze test, a sentence context is given as a prompt and participants have to respond with the first continuation that comes to their mind. However, a high cloze probability does not necessarily imply that a specific word is indeed predicted by a participant; there is a difference between cloze probability and actual prediction accuracy, as shown, for example, in the study by Brothers et al. (2015) referred to above. Although not further relevant, two additional terms that have been associated with prediction should be mentioned: Another measure of predictability is *surprisal*, which is considered to be high when predictability is low. The surprisal for a word is defined as the log ratio of the prefix probability accumulated through the words in the preceding context compared to the prefix probability after encountering it. The term was introduced by Hale (2001) to account for processing difficulties in reading; see also Boston, Hale, Kliegl, Patil, and Vasishth (2008) and Levy (2008). Yet another measure is the *entropy* of a context, defining the degree of uncertainty about its continuation. Simplifying, Pickering and Gambi (2018) describe the entropy of a context as being low when the context provides a high cloze probability for an upcoming item and, conversely, as being high when the context is a low-cloze one. The entropy, like surprisal, can be defined with respect to words but also syntactic categories or structures.

2.2.2 Underlying mechanisms

2.2.2.1 Proactive prediction versus priming

Several researchers further discriminate between underlying mechanisms and, for example, exclude the pre-activation of lexical items through priming from proactive prediction. Following DeLong et al. (2014), pre-activation through priming is not the result of high-level contextual representations (the message-level interpretation or situation model), but of associations with pre-stored linguistic units like words. Lau, Holcomb, and Kuperberg (2013) base the distinction on memory representations: Whereas prediction is seen as a process where items are predictively added to the contextual representation of the ongoing speech stream held in working memory, a related word and its features may also be activated by stored representations in the long-term memory. Critically, priming is assumed to be a passive mechanism and no costs are expected. The difference between proactive prediction and priming can be explained using the famous example of predictive processing provided in the introduction, the VW experiment conducted by Altmann and Kamide (1999). One may argue that this experiment did not assess prediction (e.g., Boland, 2005, p. 240). The fact that participants anticipated the *cake* after *The boy will eat* may show that the participants used the lexical-semantics of the verb *eat* to actively predict the upcoming object, but this finding may also be explained in terms of local priming, i.e., *eat* activates cake, the only edible object in the visual scene, through simple association. There is research showing that words prime words that are in one way or another related when the object is shown visually: Huettig and Altmann (2005) showed that words belonging to the same semantic category can prime looks to a depicted object; for example, when hearing *piano* people look at a depicted *trumpet*. This also works for more broadly semantically related words. In a study by Yee and Sedivy (2006) people looked more at the depicted *key* than at the unrelated objects when listening to *lock*. Or it occurs with shared visual features, as in Dahan and Tanenhaus (2005), who could show that people initially tended to look at the *rope* when the target was *snake*.

To systematically test the influence of priming, Kukona, Fang, Aicher, Chen, and Magnuson (2011) conducted two VW experiments. In both, the introduced character *Toby* was seen in the middle of the screen as the central fixation point surrounded by photos of four other characters. Like Altmann and Kamide, the authors used either a restrictive (4-a)/(5-a) or non-restrictive verb (4-b)/(5-b). In addition, they used two types of trials: For the first experiment (4), in competitor absent trials, the display included a picture of a *crook*, while in competitor present trials, a *policeman* was also shown. In the first experiment with active sentences, the crook is a good Patient of the verb *arrest*, whereas the policeman is a good Agent; however this thematic role is already filled by *Toby*.

(4) a. Toby arrests the crook.
    b. Toby notices the crook.

(5) a. Toby was arrested by the policeman.
    b. Toby was noticed by the policeman.

The results showed more fixations on the target Patient (crook) after restrictive verbs than after non-restrictive verbs in the anticipatory time window, lasting from the mean verb offset until 200 ms after noun onset, for the competitor absent trials, replicating the results of Altmann and Kamide (1999). In trials in which the competitor was present, participants also fixated the competitor Agent (policeman), although the Agent role was already filled, providing evidence for local priming. Nevertheless, overall anticipation of target Patients was greater than for competitor Agents. In a second experiment, the same design was employed, but this time passive sentences were used (5), thus the critical pictures became target Agents (policeman) and competitor Patients (crook). Again, in addition to an effect of verb manipulation, there was an advantage for competitors over the unrelated distractors, but the effect of local priming was reduced: The advantage of targets over competitors rapidly emerged. Based on their findings, Kukona et al. (2011) suggest “think[ing] of local priming influences on anticipatory eye movements along a continuum” (p. 35), with contexts where local priming plays a major role at one end and contexts where proactive prediction is required and results cannot be explained through priming at the other end, whereby their findings would fall in between; for ERP studies testing the influence of priming, see, e.g., Lau et al. (2013), Otten, Nieuwland, and van Berkum (2007) and Otten and van Berkum (2008). Since it is often difficult to tease apart proactive prediction and priming, especially in visual-world eye-tracking, and probably both interact, I will not discriminate between them, but subsume both under the umbrella term prediction.

2.2.2.2 Prediction by production

Researchers like Huettig (2015) assume that the mind/brain employs even more mechanisms to predict upcoming information. He proposes that at least the following are minimally required: (i) the language production system – a language user could employ fully-specified production representations to anticipate what another speaker is going to say, (ii) simple associative mechanisms, whereby different forms of priming can be involved, (iii) combinatorial mechanisms sensitive to multiple constraints and, in addition, (iv) event simulation or people's mental imagery. He uses the abbreviation *PACS* (Production-, Association-, Combinatorial-, Simulation-based prediction) to refer to this approach.

Slightly different from Huettig (2015), Pickering and Garrod (2007) have argued that the production system is used to predict upcoming input during comprehension; for an extended approach, see Pickering and Garrod (2013). Recent experimental evidence that the language production system is recruited during sentence comprehension is provided by an ERP study by Martin, Branzi, and Bar (2018). They could show that, when engaged in a secondary task that prevents inner speech through articulatory suppression, here the repetition of a certain syllable, the N400 component elicited by an unexpected gender-marked article preceding the target noun was hindered, whereas it was not when the secondary task was to produce just noise, or participants instead listened to their own voice repeating the syllable. Hence, the availability of the production system seems to be a necessary requirement for prediction during language comprehension. Another interesting finding of this experiment was that a lack of prediction did not hinder proper integration of the noun, a point discussed later in subsection 3.1.3.

Recently, a similar approach of prediction-by-production was presented by Pickering and Gambi (2018), an approach I describe here in more detail, because I will return to it several times. They propose that the comprehender first covertly imitates what is said and then derives the intention from the underlying utterance, while also taking into account background knowledge and other extralinguistic information like the visual context assumed to be shared between speaker and listener. At the same time, the comprehender may compensate for differences between herself and the speaker. The derived intention is then run through the comprehender's own production system, triggering the retrieval and building up of production representations. These representations constitute the comprehender's prediction of the speaker's upcoming utterance. Representations cover all linguistic levels, however lower levels follow higher levels as in language production (semantics > syntax > (phonological) form) and can be dependent on the time and resources available to the comprehender. In addition to the simulation route, the comprehender may rely on the less resource-intensive association route. This route corresponds to what has been described before as priming. Since spreading activation between linked representations in long-term memory is less resource-intensive, this largely automatic process may still benefit comprehension, although activation through association has been shown to decay very rapidly. Different from the prediction-by-production or simulation route, prediction-by-association would be non-optional. The optionality of prediction-by-production, the authors argue, is seen when considering prediction in other populations like L2 speakers, who show no prediction at later stages like syntax and phonology (see the review in section 4.2), or older adults. Studies on prediction in children hint at the emergence of predictive abilities at a very young age. However, as shown by Gambi, Gorrie, Pickering, and Rabagliati (2018), later stages, in particular phonology, may remain problematic for children; see section 4.3.

2.2.2.3 Bayesian inference

One branch of research in addition to psycho- and neurolinguistics that tries to capture predictive language processing is computational linguistics and neuroscience. Based on grammatical rules of the language, computational linguists try to build appropriate models (e.g., Hale, 2001) or networks (e.g., Altmann, 2002; Dienes, Altmann, & Gao, 1999; Elman, 1990). One way to formalize prediction is Bayesian inference. The Bayes' rule or theorem is a formula that allows the computation of posterior probabilities of hypotheses from prior probabilities, using the concept of conditional probability (e.g., Puga, Krzywinski, & Altman, 2015). How can this formula explain predictive language processing? Following the HMRGF by Kuperberg and Jaeger (2016), parsing hypotheses are constantly updated in an incremental fashion: As each word comes in, a prior probability distribution is shifted to a new posterior probability distribution, which in turn becomes the prior distribution for a new cycle of belief updating. The new prior probability distribution corresponds to probabilistic predictions for a new sentence structure starting with the next cycle; for evidence that the brain operates according to Bayesian principles, see, e.g., Bar (2007). In subsection 8.1.1, I will present an approach to how coreference establishment can be formalized using the Bayes' rule. Another approach is that of predictive coding, which makes use of the *prediction error*, the difference between a perceptual input and a prediction. Aitchison and Lengyel (2017) combine predictive coding and Bayesian inference in one framework, as neural responses can be described in terms of predictive coding, but this cannot account for how predictions are computed, which Bayesian inference does. The authors, however, emphasize that Bayesian inference as a computational goal can be performed by other neural algorithms and representations.

2.3 Summary of the methods and key technical terms

This chapter has introduced the method of eye-tracking during listening, specifically the visual-world eye-tracking paradigm, as a method that is ideal for measuring prediction. The method has been frequently employed in experiments with adult L1 speakers and experiments with children and adult L2 speakers. Although frequently used, there is no consensus among researchers on how to conduct and especially how to analyze VW data. Here, the growth curve analysis, a type of multilevel regression, has been described in more detail as it will be the primary analysis in the experiments to follow. This analysis enables the researcher to uncover the trajectory of an effect. Collapsing observations within a time window can distort the results by ignoring baseline and rate effects (see Barr, 2008). An effect of prediction, as it has been defined here, means a higher/increasing likelihood of looks to a target picture relative to other pictures in the visual display in a predefined time window. To measure prediction (and not integration), the critical time window should have its offset prior to the onset of the critical perceptual input, that is, before information about the target picture is encountered. Within this chapter it has further been highlighted that prediction is often used as an umbrella term and is used interchangeably with anticipation, although the latter is sometimes used to refer to even broader expectations. Some researchers further discriminate between proactive prediction, which can be associated with processing costs, and priming as a more passive and non-optional mechanism that is less resource-intensive. Since it is difficult to disentangle these mechanisms and since proactive prediction and priming probably interact, I have decided not to discriminate between them but refer to them as prediction/anticipation. There is some indication that even more mechanisms underlie prediction and that the production system and internal simulations are involved when predicting during language comprehension. Bayesian inference has been presented as a way to formalize prediction by making use of the concept of conditional probability.

Chapter 3

Theoretical background

3.1 Characteristics of child L1 and late L2 acquisition

There are many differences in how young and older children/adolescents/adults typically acquire a language. Hopp (2012, 2016), see also Grüter, Lew-Williams, and Fernald (2012), highlights the different contexts for children and late learners, that is, immersion in the case of children and, most often, formal instruction in the case of late L2 learners. Whereas young children start listening to non-segmented speech and need to detect co-occurrence probabilities between, for example, determiner-noun sequences to successfully segment the noun from the input and map it to a conceptual representation, late L2 learners already have metalinguistic knowledge about nouns and, depending on their L1, determiners and are typically exposed to written input visually pre-segmented by gaps between individual words. As a consequence, children may develop a strong link between determiners and nouns and, thus, also a strong lexical gender node. Since late L2 learners, in contrast, can map a noun directly to an existing conceptual representation, the lexical gender node is likely to be weaker. This has been taken by these authors as an explanation for the variability in the predictive/facilitative use of gender marking on Spanish and German determiners in L2 speakers (see subsection 4.2.4). Under such a view, L1/L2 differences can be traced back to lexical representations. However, there is also good reason to assume that differences in the performance of L1 and L2 speakers are the result of differences in the syntactic representation they build during processing, as described in more detail under subsection 3.2.3 in this chapter.

Within this thesis, it has not been my aim to test different L2 processing accounts, described under section 3.2, nor to argue in favor of one or another language acquisition theory, but rather to test the assumptions of the RAGE hypothesis, which predicts an L1/L2 difference in the application of prediction as a processing mechanism. Assuming that L2 speakers have a reduced ability to generate expectations, what could be the reason? To get a better idea, it appears useful to first have a look at the characteristics of child L1 and late L2 acquisition, before turning to L2 sentence processing.

3.1.1 The child learner – Starting small

Whereas the general abilities of a child grow as the child grows, the ability to learn seems to decline with age. A reason why young children in contrast to adults learn a language with ease, according to Newport (1990), may actually be because that they start small as put forward in the 'Less is More' hypothesis and the complementary 'Starting Small' hypothesis by Elman (1993). The 'Less is More' hypothesis remains agnostic to the question whether there is a special language faculty that is innate and only fully evolves early in life. However, it provides a possible explanation for the paradox that the ability to learn a language declines while general cognitive abilities increase by assuming a relationship between them. Following Newport (1990), the limited abilities of children might be advantageous, for example for the acquisition of morphology, because instead of storing complex linguistic stimuli holistically, children may only perceive and store the component parts of form and meaning. This could, under certain circumstances, reduce the number of computations needed to arrive at the correct analysis; for an implementation of how *less* can help children to learn the rules of a language, see, for example, Yang (2016). Thus, at least some constraints crucial for successful language acquisition might be non-linguistic constraints on perception and memory. While children are probably better in componential analysis, they might face difficulties with other types of stimuli that require integration and/or control over complex wholes. Elman (1993) particularly highlights the developmental change not only of the child, for example in terms of memory and attention span, but also of the child's environment, where, in the beginning, the child is typically presented with less complex sentence structures. The assumption that less could be more is mostly based on computational simulations and has been controversially discussed by Phillips and Ehrenhofer (2015); see also subsection 3.1.3.

One characteristic of children's sentence processing is a sustained problem with re-analysis, also termed the *kindergarten-path effect*. At the same time, studies on this phenomenon have found that children perform worse than adults in coordinating multiple sources of information. Trueswell et al. (1999), as one of the first studies, showed that five-year-old children, unlike adults, were less likely to revise an initial Goal interpretation in temporarily ambiguous sentences like (6-a) as shown by their eye-movement as well as their final act-out pattern.

(6) a. Put the frog on the napkin into the box. (ambiguous)
    b. Put the frog that's on the napkin into the box. (unambiguous)
    1-referent context: frog on a napkin, other animal, empty napkin, box

In unambiguous sentences like (6-b), on the other hand, the first prepositional phrase (PP) was overwhelmingly interpreted as an NP modifier, supporting the conclusion that the difficulty with (6-a) is related to the sentence being a garden-path sentence, meaning that the parser commits to a sentence interpretation that turns out to be incorrect only after later information arrives. Another, crucial finding of this study was a different use of the referential context. While the adults quickly made use of the referential context, i.e., their eye-movement pattern indicated that they interpreted the PP *on the napkin* as a modifier when two frogs were present in the visual display, the children still misinterpreted it as a Goal. Further studies could replicate these findings (Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000) and show that adultlike processing of these structures emerges around the age of eight years (Weighall, 2008).

In the English example above, the children relied on the first cue encountered, here the verb (for findings in Korean, a verb-final language, see Choi & Trueswell, 2010). The verb *put* requires a Goal argument, so there is a strong lexical bias to interpret the following PP as a destination. Adults have been shown to override this bias when the referential context supports the modifier interpretation. Note though that the results from Trueswell and colleagues can also be explained in terms of a preference for the simpler syntactic structure in children as argued by Traxler (2002). Snedeker and Trueswell (2004) could show that English children at the age of four to five years exclusively relied on a highly reliable cue like the lexical bias of verbs as visible from their act-out pattern, although their eye-movements revealed emerging sensitivity to referential context; for similar child/adult differences, see Kidd, Stewart, and Serratrice (2011). Findings like these have been taken as an indication that children have available a probabilistic multiple-cue comprehension system from the start, where the most reliable cue takes precedence, also known as constraint-based learning, and they do not only use single cues (Snedeker & Trueswell, 2004). This assumption is similar to the Competition Model described in the next subsection. Children's difficulty with re-analysis may also account for other difficulties, like the processing of passives (Huang, Zheng, Meng, & Snedeker, 2013), and seems to persist up to the age of twelve as shown in a reading study by Traxler (2002). Novick, Trueswell, and Thompson-Schill (2005) report that the prefrontal cortex, including the left-inferior frontal gyrus known to be involved in cognitive control processes, is among the last neuroanatomical regions to mature and that underdevelopment affects sentence comprehension, as has also been shown for patients with lesions in these areas.

One observation described above, the discrepancy between eye-movement data and final decisions or between online and offline data, is another characteristic of sentence processing in children. In an online picture selection task that required the participants to indicate their final decision, Sekerina, Stromswold, and Hestvik (2004) investigated adults' and children's interpretation of ambiguous pronouns in English. Unlike in (7-a), the short-distance pronoun in the example from Sekerina et al. that is shown in (7-b) can refer either to the sentence-internal referent, the boy, or to a sentence-external referent, another male person. The participants saw two pictures displayed next to each other, one showing the event where the box is placed behind the boy (sentence-internal) and another where the box is placed behind a man (sentence-external). The authors observed competition between a sentence-internal and sentence-external referent interpretation in the eye-movement pattern of the adults, and, although the sentence-internal referent was the more accessible referent, still 20% of the final choices were for the sentence-external referent interpretation. The four- to seven-year-old children, in contrast, preferred the sentence-internal referent offline (97% choices), although their eye-movements showed implicit awareness of the referential ambiguity. The online data also showed that the children were slower than the adults in recognizing the ambiguity; it took them around 1000 ms longer. The authors concluded that children need more time to access discourse reference and integrate discourse information.

(7) a. The boy has placed the box behind himself.
    b. The boy has placed the box behind him.

All the phenomena described have also contributed to the general assumption that adults' first interpretation might be children's only interpretation (e.g., Phillips & Ehrenhofer, 2015).

To summarize, young children may have an advantage in learning a language due to the fact that their limited cognitive abilities force them to store linguistic stimuli not holistically but analytically, which, in turn, makes certain analyses more efficient. On the other hand, at least up to the age of eight years but for some linguistic structures probably beyond, children have difficulties in revising initial analyses. Since experiments show differences between offline and online performance, conclusions about sensitivity to information sources in children are better drawn from the latter.

3.1.2 The late second language learner – Starting with experience

Above, the ease with which children acquire parts of the language has been explained in terms of maturational constraints as proposed by the 'Less is More' hypothesis. There is an ongoing debate on whether and how the AoA affects the L2 outcome; see a recent discussion in *Bilingualism: Language and Cognition* (Abutalebi & Clahsen, 2018). Newport (1990), see also Johnson and Newport (1989), found that, whereas the age of arrival in the U.S. and the performance on a test on English morphology and syntax were highly correlated in a group of Chinese and Korean L1 speakers who arrived before puberty, this was not the case for those who arrived later. More recently, Hartshorne, Tenenbaum, and Pinker (2018) conducted a survey with more than 600 thousand native and non-native English speakers and discovered that around the age of 17 years, grammar-learning abilities start to decline. That is much later than previously assumed. Hence, one may assume that there is a critical or sensitive period or, related to the linguistic domain, even multiple periods for L2 acquisition (see also the introduction in Slabakova, 2008). However, AoA is one factor. A further factor that is different in child L1 and later L2 acquisition, as mentioned before, is the type of exposure and/or context of learning. Many L2 speakers learn the L2 through formal instruction, for example in a classroom environment. Thus, later L2 acquisition exhibits a diversity in language learning settings, but also the learner characteristics and circumstances are diverse. Social and individual factors can influence the success of L2 acquisition; Saville-Troike and Barto (2012) further mention a learner's aptitude and the reason or motivation for learning the language.

Intriguing when comparing child L1 and late L2 acquisition is that sometimes different acquisition sequences are observed as, for example, reported by Clahsen and Muysken (1986) for verb placement in German. German has two dominant word orders, SVO in main clauses and SOV in subordinate clauses. While German-speaking children were found to start with a preference for verb-final placement, which typically is assumed to be the basic or underlying word order, adult L2 learners with Italian or Spanish as L1, who learned German through immersion, first tended to place the verb after the subject. The authors made similar observations in utterances of Turkish L1-German L2 speakers, although Turkish also places the lexical verb at the end of subordinate clauses, hence the data cannot be explained in terms of transfer from the L1. The reason why L2 learners of German may arise at a different assumption about basic word order than children, according to Meisel (2013), could be that they rely more on surface properties. If the surface word order in many syntactic contexts is SVO, they may use inductive learning strategies to formulate an incorrect hypothesis. Note that the author assumes that there are universals of linguistic structure and that structures are derived.

Differences between child L1 and late L2 learners have also been observed for morphologically complex words. Children are reported to overregularize and produce plural forms like **mouses* instead of *mice*, although they may know the correct form; even in German, where the *-s* plural is far less frequent, overapplication of *-s* has been reported. Similar observations have been made for regular and irregular past tense forms (e.g., Clahsen, 1999; Marcus et al., 1992). Marcus et al. (1992) suggest that children make use of an affixation rule and that retrieval of an irregular form can block that rule, however due to children's limited memory traces, retrieval sometimes fails. Interestingly, L2 speakers tend to memorize not only exceptions like *went* as the past form of *to go*, but also regularly inflected verbs like *walked* as shown in experiments using, for example, the masked priming technique (Felser & Clahsen, 2009).[2] According to dual-mechanism approaches, there are two systems, the memorized mental lexicon and a computational mental grammar. Irregular (non-default) forms are stored in the former system, while regular (default) forms are computed by rule application in the latter. Ullman (2001a, 2001b, 2005) terms these two systems the *declarative memory system* and the *procedural memory system*. Psycho- and neurolinguistic studies support that L2 speakers rely more on the declarative memory system than L1 speakers and store inflected forms holistically, at least initially. They may also learn explicit rules in the declarative memory system (Ullman, 2001a). However, the lexical-declarative system cannot take over

[2]Here, participants are presented with a prime word (*walked*) for a very short time, so they are not consciously aware of it, before being shown a target word (*walk*). The task is to indicate whether the target word is an existing word (lexical decision). Reaction times are compared to those for unrelated prime words and/or identical words as primes.

all functions served by the grammatical-procedural system, so reliance on just this system should not lead to nativelike proficiency at all levels of grammar.

Within the Competition Model (e.g., MacWhinney, 2001; MacWhinney, Bates, & Kliegl, 1984), L1 and L2 acquisition is seen as a constructive, data-driven process that relies on universals of cognitive structure. The model was designed to quantify the ways in which distributional properties of the input control the learning and processing of a language. Its underlying assumption is that the language learner needs to detect certain cues and their reliability and availability, which determine their strength. The more reliable a cue, the earlier it should be acquired. Due to the availability of a well-organized neurolinguistic system, L2 learners are expected to show transfer from their L1; however, a consequence of learning would be the tightening of within-language links in contrast to between-language links. Often L2 learners also have to learn new conceptual or discourse distinctions. MacWhinney (2001) mentions the acquisition of verbal conjugation in Hungarian. There are 13 conditions in Hungarian that control whether the transitive or intransitive verb conjugation is used, whereby the choice is influenced by transitivity, definiteness and referential disambiguation, but there is no single principle. Whereas the child would learn this by generalizing from a rich database of collocations and phrases, similar to what has been stated for determiner-noun pairs, the L2 learner would need to focus on contrastive structures in which cues are competing, thus a different outcome is expected, with the L2 speaker being prone to errors.

To summarize, L2 acquisition is different from child L1 acquisition, and an L2 learner's target grammar and/or processing might be different from that of a native speaker, a point I will turn to in section 3.2. A possible explanation is the different starting point, with the later L2 learner having access to an L1 and general learning strategies like inductive reasoning. More or differently than in L1 acquisition, L2 acquisition is influenced by a range of factors like AoA, the context of learning (immersion, instruction) and other individual differences.

3.1.3 Sentence processing in language learners

As will be shown in chapter 4, children learning their L1 have been found to predict upcoming information from very early on. This is in contrast to what has been reported for adult L2 speakers. Phillips and Ehrenhofer (2015) speculate that late learners of a language might be held back by what they have learned at earlier stages of learning. Since they are better in (re-)analysis (though see for example Jacob and Felser (2016) and Pozzan and Trueswell (2016) for evidence against such a claim), but not yet good in prediction, this might let them maintain sentence processing routines, so they will continue to wait for upcoming information to appear. In section 2.2, an ERP experiment by Martin et al. (2018) was described that found that L1 speakers who did not predict still had no problems in integrating arriving information, indicating that prediction is not a necessary prerequisite for successful sentence interpretation (see also Huettig &

Mani, 2015). Hence, even with increasing proficiency, L2 speakers may be 'blocked' and do not make use of predictive cues. Young children acquiring their L1, in contrast, may become more efficient in predicting, the more proficient they become. Phillips and Ehrenhofer argue that prediction helps learners to test hypotheses about a language and can be particularly helpful in overcoming overgeneralizations. The importance of prediction for language learning, as mentioned in the introduction of this thesis, has also been highlighted by other researchers (e.g., Chang et al., 2013).

Phillips and Ehrenhofer (2015) also remark that current research lacks experiments on more complex predictive cues with children. It remains unclear for the time being how, for example, children integrate information sources from different linguistic domains to form a prediction. A difficulty with the real-time integration of cues may be found in child L1 and adult L2 speakers alike; see also Clahsen and Felser (2006a, p. 4). Although children may benefit from their limited resources in some domains, they certainly do not in others.

3.2 L2 processing

This section focuses on sentence processing in L2 speakers. There are several psycholinguistic approaches that have tried to explain the L1/L2 differences observed here. Some explain these in terms of resource deficits in L2 processing, including memory capacity, lexical decoding ability and processing speed (e.g., Hopp, 2010; McDonald, 2006). An extension to these limited-capacity models is the Lexical Bottleneck Hypothesis put forward by Hopp (2018). More fundamental differences are proposed by the Shallow Structure Hypothesis (Clahsen & Felser, 2006a, 2006b, 2006c, 2018) and the Interface Hypothesis (e.g., Sorace, 2011; Sorace & Serratrice, 2009). More recently, Cunnings (2017) claimed that L1/L2 differences result from a susceptibility to interference during memory retrieval operations. Another approach and the one under investigation in this thesis is that L2 speakers differ from L1 speakers in the way they predict upcoming information (Grüter et al., 2014; Grüter, Rohde, & Schafer, 2017; Kaan, Dallas, & Wijnen, 2010). In the following, I try to show that a difference in the application of prediction during processing alone cannot explain the whole range of differences observed, but that prediction might be differently affected. I will consider general factors that may affect prediction as proposed by accounts arguing for L1/L2 similarity, but also limits that are selective and more specific for L2 processing as proposed by accounts arguing for more substantial L1/L2 differences.

3.2.1 The RAGE hypothesis

Research in the area of predictive processing has shown that prediction poses a challenge for L2 speakers even at higher proficiency levels. From the results of their ERP experiment testing the prediction of an upcoming word and its phonological form, see subsection 4.2.2, Martin et al. (2013) concluded that L2 speakers' active lexical prediction is limited or less efficient compared to L1 speakers' and that this might be the result of lower and less accurate linguistic processing stages. Lew-Williams and Fernald (2010) and Grüter et al. (2012), who tested the use of gender information on Spanish determiners, found no facilitation effect for L2 speakers, see subsection 4.2.4, but they did for two- to three-year-old Spanish-speaking children (Lew-Williams & Fernald, 2007). Findings like these, together with their data from a story continuation task testing expectations at the discourse-level, see subsection 4.2.6, led Grüter et al. (2014, 2017) to formulate the RAGE hypothesis, which states that non-native speakers have a (R)educed (A)bility to (G)enerate (E)xpectations.

> "[N]on-native speakers have a reduced ability to generate expectations, that is, reduced ability to engage in proactive processing, while their abilities in information integration, or reactive processing, may be more closely aligned with those of native speakers." (Grüter et al., 2014, p. 189)

As more and more studies on the topic have been published, it has become clear that the RAGE hypothesis needs refinement. Grüter et al. (2014, 2017) mention further factors that probably influence predictive processing, like language proficiency and L1-L2 typologies, but also linguistic domains, an aspect addressed in subsection 3.2.3. In Grüter, Lau, and Ling (2018a), the relative weighting of cues is mentioned and the authors, in line with Clahsen and Felser (2018), assume the differences between L2 and L1 speakers to be gradual and not categorical. The current thesis follows up on these more recent changes by testing highly proficient L2 speakers on their use of different information sources that can function as cues to sentence interpretation in both the L1 and L2.

If it is not prediction as a processing mechanism that is different between L1 and L2 speakers, but prediction being affected differently, what could be the reason for this? This question will be addressed in the following two subsections.

3.2.2 Individual differences

Kaan (2014) argues that it is not necessarily the case that L2 speakers differ from L1 speakers in terms of different processing mechanisms; the observed differences between L1 and L2 processing rather result from individual differences that can also be responsible for differences in L1 speakers. One factor that varies between language users is the frequency information they have stored. L1 speakers could, for example, use verb argument structure biases to predict an upcoming argument. However, the frequency with which language users are exposed to such preferences is subject to individual differences, depending on the input they receive. People being exposed to different language varieties may differ from each other. For L2 learners in particular, the input is a different one when being formally instructed in a classroom environment compared to learning through immersion, or when frequently listening to other learners.

Thus, the type of exposure may influence prediction. L2 learners typically differ from L1 speakers in the amount of exposure, receiving less input in the L2.

This relates to another factor, namely that of language similarity and experience. One study that provides evidence against the assumption that L2 speakers' predictive processing abilities are generally limited comes from Foucart, Martin, Moreno, and Costa (2014), who revised the claim made in Martin et al. (2013). Actually, Martin et al. tested a phenomenon that does not exist in the L1 of their non-native group, namely phonological agreement. When testing late learners of Spanish with French as L1, a language that also marks gender on pre-nominal determiners, the L2 group showed the same biphasic pattern as the Spanish L1 group and the Catalan-Spanish early bilinguals they were compared to. In the following literature review, language similarity will be considered as a factor that may influence prediction in that the absence or different realization of a cue in the L1 may hinder prediction in the L2. A further factor highlighted by Kaan (2014) that might lead to differences is competing information. Language users have been shown to activate their L1 when processing their L2. This additional information activation might impede or slow down predictions during L2 processing, thus resulting in L1/L2 differences. For example, Wu and Thierry (2010) could show that proficient L2 speakers of English implicitly activated their L1 Chinese while reading or listening to English. However, the two studies considering this (Chambers & Cooke, 2009; Ito, Pickering, & Corley, 2018) found no evidence of parallel activation or at least no influence of L1 activation on L2 speakers' predictive abilities. Chambers and Cooke (2009) only report that L2 proficiency affected anticipation.

Another factor mentioned when prediction was defined was the availability of cognitive resources. These and cognitive control abilities may also play a role in prediction (see Kaan, 2014), particularly in L2 speakers. Huettig (2015) points to the role of working memory capacity as more recently demonstrated in Huettig and Janse (2016) for L1 speakers between the ages of 32 and 77 years. In visual-world eye-tracking experiments, the depicted objects activate a visual representation, which cascades to higher levels of representation. Following Huettig (2015), the knowledge associated with these objects as well as linguistic knowledge like the representation of their gender are connected to the object's location within working memory. In addition, a lack of fast and efficient engagement of production processes might hinder certain participant groups from predicting upcoming input (Martin et al., 2018, p. 7f.). However in a study by Ito, Corley, and Pickering (2018), see subsection 4.2.1, the authors reported no L1/L2 differences for prediction based on verb semantics tested in a cognitive load condition. Kaan (2014) further mentions as factors that potentially influence predictive processing task-induced processes and strategies, thus the way an experiment is designed or conducted could affect processing, and L1 and L2 speakers might be sensitive to this to varying degrees, as well as the motivation people have. Lim and Christianson (2013), for example, found that reading for comprehension and reading for translation affected how the L2 was processed. Especially low-proficient L2 speakers seemed to allocate more attention to structural information rather than semantic information or world knowledge when the reading goal was translation.

A further factor that has been shown to result in differences between L1 speakers and that might play a role in predictive processing in an L2, according to Kaan (2014), is the accuracy and consistency of lexical information. If the lexical representation is not accurate and/or is inconsistent, retrieving it during processing and, moreover, using it to predict is far more effortful. A study by Hopp (2013), see subsection 4.2.4, found that L2 speakers who consistently assigned the correct gender in production, were also more likely to use it in an anticipatory fashion. Moreover, faster access to and use of lexical items led to a stronger effect (see also Hopp, 2016). Hopp (2016) could show that, when variability was created in the input of L1 speakers, they stopped using gender predictively, demonstrating that variability of gender assignment modulates prediction not only in L2 speakers. This aspect fits in an approach known as the Lexical Bottleneck Hypothesis (see most recently Hopp, 2018). According to this hypothesis, weaker lexical links together with more diffuse lexical activation and representation may account for L1/L2 differences. These are expected to affect syntactic processing, which follows lexical processing. Hopp particularly highlights the role of non-selective lexical access. As discussed above, this implies the activation of the L1 during L2 processing. In line with Kaan (2014), this approach does not assume fundamental differences between L1 and L2 speakers but rather that reduced prediction in L2 speakers can be attributed to characteristics of the bilingual mental lexicon.

3.2.3 Selective differences

The idea that prediction might be selectively limited is not new. Kaan et al. (2010) already suggested that L2 speakers do not actively predict syntactic elements during online processing, transferring assumptions of the Shallow Structure Hypothesis (Clahsen & Felser, 2006a, 2006b) to processing mechanisms. However, the authors assumed that with increasing proficiency also the ability to predict might become nativelike. In the following, the Shallow Structure Hypothesis (SSH) and the Interface Hypothesis (IH) will be presented as two accounts that motivate why prediction based on grammatical cues or at the discourse-level might be different between L1 and even highly proficient L2 speakers.

The SSH proposes a different use of information sources in L1 and L2 speakers. The SSH put forward by Clahsen and Felser (2006a, 2006b, 2006c) builds up on existing approaches in psycholinguistics developed to account for findings from L1 processing that suggest the availability of different processing routes (e.g., Ferreira, Bailey, & Ferraro, 2002).[3] One such route is deep and involves full parsing, i.e., a

[3]The Good Enough approach (Ferreira et al., 2002) argues for two processing routes: a deep and a shallow route. To obtain the meaning of a sentence a parser may not build a complete and detailed representation. Thus, a sentence like *The dog was bitten by the man* can be interpreted as a man was bitten by a dog, which better aligns with common world knowledge, employing the shallow route that relies more on semantic heuristics and not on syntactic algorithms.

complete specified syntactic representation for an upcoming string of words, while the other is shallow and based on lexical-semantic information, associative patterns and other surface cues to interpretation; both processing routes can also operate in parallel. Importantly, the assumption is that both L1 and L2 speakers have the same parsing mechanisms available. However, L1 and L2 processing systems may differ in the way they make use of different information sources – in terms of grammatical constraints being absent, represented incorrectly or differently weighted – and thus different routes are employed, and/or the L2 processing system relies more strongly on non-grammatical than on grammatical information. In this way the SSH discriminates between the knowledge or grammar and the parser. The difference from other accounts is that differences between L1 and L2 speakers in reliance on these two routes are not assumed to be only quantitative.

> "Adult learners are guided by lexical-semantic cues during parsing in the same way as native speakers, but less so by syntactic information. We suggest that the observed L1/L2 differences can be explained by assuming that the syntactic representations adult L2 learners compute during comprehension are shallower and less detailed than those of native speakers." (Clahsen & Felser, 2006a, p. 3)

The authors assume the differences between L1 and L2 processing to be gradual, a point they particularly highlight in Clahsen and Felser (2018), where they try to clarify prior misinterpretations of the SSH. In the original article, Clahsen and Felser (2006a) stated that L2 speakers would show greater reliance on lexical-semantic, pragmatic and surface information, e.g., when processing non-local dependencies, instead of syntactic information, leading to shallower and less detailed syntactic representations. This, however, should not imply that L2 speakers' processing is limited to shallow parsing. Moreover, the SSH does not rule out L1 transfer or an effect of cognitive resources, but questions whether these can fully explain the differences observed between L1 and L2 processing. Crucial for the question here, namely whether there is a difference in the use of information sources for prediction, is the assumption that grammatical constraints may be less robust in L2 processing.

Compatible with the assumptions of the SSH is the assumption that L2 speakers have difficulties with the integration of different information sources. This is the core claim of the IH, which, however, further assumes difficulties at interfaces, in particular the syntax-discourse interface. Sorace and Serratrice (2009) discriminate between two interfaces: the internal syntax-semantics interface and the external syntax-discourse interface. The syntax-semantics interface involves formal features and operations within syntax and logical form, for example in the case of specific and generic NPs as shown in (8) for English and (9) for Italian. Violations at this interface are assumed to result in ungrammaticality.

(8) a. ∅ Sharks are dangerous animals. (generic)

b. **The** sharks at the aquarium are rather small. (specific)

(9) a. **Gli** squali sono animali pericolosi. (generic)
the sharks are animals dangerous
'Sharks are dangerous animals.'

b. **Gli** squali all' acquario sono piuttosto piccoli. (specific)
The sharks at the aquarium are rather small

The syntax-discourse interface involves pragmatic conditions that determine appropriateness in context, for example the different distribution of subject pronouns for English (10) and Italian (11). Here, violations would be not categorical but gradient, meaning that certain forms are more or less appropriate in the current discourse context.[4]

(10) a. While John is eating, **he** (John) is talking on the phone. (same topic)

b. While John is eating, **he** (Paul) is talking on the phone. (different topic)

(11) a. Mentre Gianni mangia, ∅ (Gianni) parla al telefono. (same topic)
while Gianni eats talks on the phone
'While Gianni eats, he (Gianni) talks on the phone.'

b. Mentre Gianni mangia, **lui** (Paolo) parla al telefono. (different topic)
while Gianni eats he talks on the phone
'While Gianni eats, he (Paolo) talks on the phone.'

The IH tries to explain why even near-native L2 speakers exhibit persisting optionality at the syntax-discourse interface. Reasons mentioned are the underspecification of interpretable features and cross-linguistic influence at the level of grammatical representation. For example, the overextension of overt subject pronouns in null-subject languages by English L1 speakers might be the result of underspecification due to the absence of a similar condition in English in the same syntactic context. Alternatively, processing limitations might let L2 speakers fall back on a default strategy, in this case overextension of the scope of overt subject pronouns.

> "[T]he IH predicts that both syntactic and pragmatic conditions are acquirable, but the integration of syntactic and pragmatic conditions remains less than optimally efficient and gives rise to optionality." (Sorace, 2011, p. 26)

According to Sorace and Serratrice (2009) and Sorace (2011), processing limitations better account for findings at external interfaces that are resource intensive rather than (only) cross-linguistic influence; moreover, developmental effects in monolingual children have been observed at external interfaces (Sorace, 2011, p. 5). Cross-linguistic factors, on the other hand, have been found to affect structures at the interface of formal features. Another influencing factor might be the input, its quantity and quality, as also discussed in the prior subsection, which may include the parallel activation of languages. A further factor that may explain bilinguals' inefficiency at computing the syntax-pragmatics mapping is the allocation

[4]The examples (8)–(10) follow the examples given by Sorace and Serratrice (2009, p. 204).

of resources rather than resource limitations, i.e., the ability to flexibly direct attentional resources as a function of the complexity of the incoming input.

3.3 Summary of the theoretical background

This chapter started with an overview of the characteristics of child L1 and late L2 acquisition. Although theories differ in their argumentation regarding the origin of the differences observed, they agree in some respects: (i) Children master the morphological system of the target language relatively early, whereas they demonstrate, at least initially, difficulty with more complex cues and/or the integration of multiple information sources and show a persistent difficulty with the revision of a first structural commitment; (ii) Late L2 speakers already come to the learning of the target language with prior knowledge and, at least initially, demonstrate difficulties with morphology and syntax. Processing research further shows that they also have difficulty with interface phenomena, particularly phenomena including the discourse-level. Phillips and Ehrenhofer (2015) attempt to link current findings in acquisition and processing research to find out more about how learning proceeds. Although the focus here is not on learning, the authors' approach is interesting in regard to the question why L1 and L2 speakers should differ in their use of prediction during processing and why it is interesting to also consider the child L1 speaker. The authors try to provide an explanation for the contrasting patterns of child L1 and late L2 learners, where children despite their limited cognitive resources in comparison to adults start to predict from very early on, arguing that children may use prediction as a hypothesis testing device, which may let them outperform the adult L2 speaker in the long run. Note, however, that this approach is very speculative.

The next section focused on L2 sentence processing. I have grouped the L2 processing accounts discussed into accounts that assume that L1 processing and L2 processing are basically the same and L2 processing accounts that assume more substantial L1/L2 differences, most notably the SSH and IH. I have highlighted the RAGE hypothesis as an approach that tries to explain L1/L2 processing differences in terms of a different application of prediction. Since more recent empirical findings have indicated that L2 speakers' prediction may not be generally reduced (see section 4.2 for a review), it is asked why prediction could be differently affected in L1 and L2 processing. I hypothesize that based on the assumptions of the SSH, L2 speakers may rely less on grammatical information than L2 speakers, not only in sentence interpretation but also when predicting upcoming input. The IH may explain why previous studies have found L1/L2 differences for expectations at the discourse-level. Following the IH, L1/L2 differences are the result of a problem with the syntax-discourse mapping in L2 speakers. Differently, the problem may not be due to difficulties at the discourse-level but rather with the real-time integration of information sources, meaning that also when no discourse-level information is involved, L1 and L2 speakers may differ from each other.

In sum, the RAGE hypothesis predicts that L2 speakers differ from L1 speakers whenever prediction is involved. However, L2 speakers' predictive abilities appear to be affected by a range of factors, as already pointed out by the hypothesis' proponents. These could be individual factors that can also affect prediction in L1 speakers, like the availability of cognitive resources, or more L2-specific individual factors, like L2 proficiency and L1-L2 similarity. Hence, to be able to argue for selective limits, another possibility for why L2 prediction has not been found to be reduced consistently, such factors should be controlled for as well as possible.

Chapter 4

Prediction review

4.1 Prediction in L1 speakers

There is plenty of evidence that adult L1 speakers can use multiple sources of information to predict upcoming information during language processing. Previously discussed examples have shown that, for instance, the lexical constraint of the context or the lexical-semantics of verbs trigger an expectation towards a specific word or feature of a word, for example being edible. Mani and Huettig (2012) found that already at the age of two years German toddlers were able to anticipate an upcoming object in a sentence like (12-a).

(12) a. Der Junge isst den großen Kuchen.
The boy eats the big cake

b. Der Junge sieht den großen Kuchen.
The boy sees the big cake

There is robust experimental evidence demonstrating that children acquiring their L1 use the selectional restrictions of verbs to anticipate a respective target object, as shown for two- and eight-year-old children in less complex eye-tracking designs that displayed only two pictures (Mani & Huettig, 2012, 2014), and for ten- to eleven-year-old children in a design with four objects including the target (Nation, Marshall, & Altmann, 2003); see also Gambi, Pickering, and Rabagliati (2016).

This chapter will show that the level of prediction is not restricted to a word and its features. Native speakers have been shown to even use different sources of information at the same time to predict. Henry, Hopp, and Jackson (2017), for example, found that cues were used additively by German L1 speakers. In their study, which included three VW experiments, the effects of verb semantics, case marking and prosodic information on predicting an upcoming thematic role, either an Agent or Patient, were tested. The authors found that the participants adapted to the utility of a cue depending on its availability. Example stimuli are shown in (13) for the manipulation through case marking and (14) shows case marking together with prosody (examples and translations are taken from Henry et al., 2017, p. 1233). Not investigated but mentioned by the authors was the reliability of a cue defined as the percentage of the time that the use of a cue would lead to an accurate interpretation. Hence, one may conclude that the use of different cues, as demonstrated by Henry et al., also depends on a cue's availability and reliability.

(13) Case:

a. Der Hahn frisst gleich die Blume.
[the chicken]-*Nom* eats soon [the flower]-*Acc*
'The chicken will soon eat the flower.'

b. Den Hahn frisst gleich der Fuchs.
[the chicken]-*Acc* eats soon [the fox]-*Nom*
'The fox will soon eat the chicken.'

(14) Case + Prosody:

a. Der Hahn frisst gleich DIE BLUME.
[the chicken]-*Nom* eats soon [the flower]-*Acc*
'The chicken will soon eat the flower.'

b. DEN HAHN frisst gleich der Fuchs.
[the chicken]-*Acc* eats soon [the fox]-*Nom*
'The fox will soon eat the chicken.'

In the following review, I will focus on studies that found an effect in advance of the perceptual bottom-up input, because this will serve as the primary criterion for prediction throughout. Note that for ERP and eye-tracking experiments an effect was defined as an ERP response or eye-movements towards a target picture prior to the onset of a critical stimulus, in the examples (13) and (14) above anticipatory eye-movements towards the picture of the flower or the fox in the adverb region *gleich* (soon). I have tried to assign the studies reviewed to either a particular source of information or a level of prediction.

4.1.1 Prediction of words and features based on prior context

The first example above, repeated in (15), from an ERP experiment by DeLong et al. (2005) showed that contextual information can already lead to facilitation before the onset of the target word.

(15) The day was breezy so the boy went outside to fly ...

a. a kite

b. an airplane

In an offline cloze test, the researchers first assessed the most likely upcoming word, which in this example would have been *a kite*. In half of the sentences they used a less likely continuation, in the example here *an airplane*, making use of the phonological rule that the English indefinite article *a* changes to *an* when followed by a vowel. DeLong et al. found that the N400 amplitude decreased the higher the cloze probability of a target noun was and, more importantly, that the effect was already visible in the N400 time window for the article, demonstrating that the phonological form of the final noun had been anticipated. More evidence for the prediction of the phonological form of a word comes from a VW study by Ito, Pickering, and Corley (2018), where they found anticipation of a phonologically related word that already emerged 500 ms before the onset of the target word (see also Martin et al. (2013) for a very similar ERP experiment; both of these experiments are described in more detail in subsection 4.2.2, as both also tested adult L2 speakers). Prediction of an upcoming word was also found in German sign language by Hosemann, Herrmann, Steinbach, Bornkessel-Schlesewsky, and Schlesewsky (2013). Sign languages are natural languages but use a different modality, i.e., manual-visual. Moreover, sign languages, unlike spoken languages, are not strictly sequential but can simultaneously use different articulators, like the phonological parameters handshape, orientation, location and movement, in the production of a lexical sign. Hosemann et al. found that, when presented with an unexpected continuation, the signers showed an N400-late positivity pattern starting already in the transition phase prior to the onset of the target sign, here the verb. This finding demonstrates that also sensory-motor information about the realization of a word or sign can be anticipated.

In a VW experiment and an ERP experiment, Rommers, Meyer, Praamstra, and Huettig (2013) could show that language users also predict the perceptual attributes of an object, here its shape. While listening to a Dutch sentence like (16), participants were presented with a four-object display showing three unrelated distractor objects and a critical object 500 ms prior to the onset of the target word.[5]

(16) In 1969 Neil Armstrong als eerste mens voet op de ...
in 1969 Neil Armstrong as first man set foot on the ...
'In 1969 Neil Armstrong was the first man to set foot on the ... '

a. maan ('moon')

b. tomaat ('tomatoe')

c. rijst ('rice')

The critical object was either the target object (moon), a shape competitor (tomato) or an unrelated control object (rice). Before information about the shape of the spoken target word was available through the input, participants showed more fixations on the target object and the shape competitor than on the control object. In the corresponding ERP experiment, the N400 amplitude in response to the final word was significantly attenuated in the shape-related condition (16-b) compared to the unrelated one (16-c).

The next experiments showed that language comprehenders can further anticipate a noun's gender feature. Wicha, Moreno, and Kutas (2004) investigated how the brain reacts to expected versus un-

[5] In (16) and throughout the text I have added the word-by-word translation to the examples if it was not provided by the authors.

expected articles, as well as nouns that agree or disagree with the gender of the preceding article or are semantically anomalous based on previous context in Spanish. The participants read the sentences word-by-word, while ERPs were recorded. For the preceding article, the authors report an enhanced, widely distributed positivity for gender-mismatching articles, appearing 500–700 ms after article onset, which they call a gender expectancy effect. Using a slightly different experimental design without overt gender agreement violations, van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005) also tested whether contextual information leads to the prediction of a specific noun, in this experiment looking at adjective-noun agreement in spoken Dutch. The materials consisted of two-sentence mini-stories like in (17). In the prediction-consistent condition (17-a), a neuter-gender noun was used, which has no overtly realized inflectional suffix, while singular common-gender was used in the condition where the adjective was inconsistent with the discourse-predictable noun (17-b).

(17) De inbreker had geen enkele moeite de geheime familiekluis te vinden.
the burglar had no effort the secret family safe to find
'The burglar had no trouble locating the secret family safe.'

a. Deze bevond zich natuurlijk achter een groot-[*neut.*] maar onopvallend schilderij-[*neut.*].
this situated itself of course behind a big-[*neut.*] but unobtrusive painting-[*neut.*]
'Of course, it was situated behind a big but unobtrusive painting.'

b. Deze bevond zich natuurlijk achter een groote-[*com.*] maar onopvallend boekenkast-[*com.*].
this situated itself of course behind a big-[*com.*] but unobtrusive bookcase-[*com.*]
'Of course, it was situated behind a big but unobtrusive bookcase.'

The results showed an N400 effect after noun onset, which was preceded by a positive deflection after adjective inflection onset for prediction-inconsistent inflections (17-b). To see whether this deflection was not just an artefact, the authors conducted a control experiment without the prediction supporting wider discourse. In the control experiment no positive deflection was visible for the same adjective inflections. An additional self-paced reading experiment, using a subset of the slightly modified materials (the region after the critical adjective was extended, e.g., *was situated behind a big but rather unobtrusive painting*), lent further support to the prediction of the noun's gender feature, with a slow-down in reading starting at the third word after the prediction-inconsistent gender-inflected adjective, i.e., one word before the critical noun; see also Otten et al. (2007) and, for written Dutch, Otten and van Berkum (2008). Foucart, Ruiz-Tada, and Costa (2015) used a design where they controlled for overlapping effects between a gender marked article and noun in speech processing and replicated the gender-mismatch effect in Spanish (for more details, see subsection 4.2.2).

The studies cited above have in common that a specific lexical item was predicted based on the previous context. Szewczyk and Schriefers (2013) found that language users can also predict semantically defined classes of words, in their particular study information about the animacy of the direct object noun. In Polish, the language investigated, grammatical gender has four classes in the singular: feminine, neuter, masculine-animate and masculine-inanimate. The authors constructed short stories ending with either a high or a low constrained sentence including a masculine-animate or a masculine-inanimate direct object. What kind of object was expected and to which degree was assessed with a cloze test. Each object was preceded by a modifier, an adjective or quantifier, that agreed with the noun in case and animacy. Prenominal modifiers that were bias-incongruent elicited a greater negativity relative to bias-congruent modifiers irrespective of predictability, i.e., whether participants were biased towards a specific word appearing in the high constrained sentences or, more general, expected a word belonging to a semantic category in the low constrained sentences. An overview of studies, including those described, that have tested the use of previous context to predict is shown in Table 4.1.

Study	**Level of prediction**
DeLong et al. (2005); Martin et al. (2013); Ito, Pickering, and Corley (2018)	phonological form of a word
Hosemann et al. (2013)	phonological form of a sign
Rommers et al. (2013)	visual feature of a noun
Wicha et al. (2004); van Berkum et al. (2005); Otten et al. (2007); Otten and van Berkum (2008, Exp.1B); Foucart et al. (2014); Foucart et al. (2015); Martin et al. (2018)	gender feature of a noun
Szewczyk and Schriefers (2013)	semantic feature of a noun

Table 4.1: Overview of experiments testing the use of prior context to predict a word and/or a particular feature

4.1.2 Prediction based on extralinguistic information

The study by Rommers et al. (2013) above showed that also visual features of a noun can be part of prediction. The next examples will show that non- or extralinguistic information can be used for predicting upcoming linguistic input as well. Knoeferle, Crocker, Scheepers, and Pickering (2005) demonstrated that role information provided by the visual scene can affect thematic role assignment. In their VW experiment, participants were presented with temporarily ambiguous German sentences as in (18).

(18) a. Die Prinzessin wäscht offensichtlich den Pirat.
The princess-*subject/object* washes apparently the pirate-*object*

b. Die Prinzessin malt offensichtlich der Fechter.
The princess-*subject/object* paints apparently the fencer-*subject*

The concurrent visual scene depicted the respective Agent-action-Patient events, for the sentences in (18) the princess in the center was facing the pirate to her left and showed the washing-event, while the fencer to her right was facing her and showed the painting-event. The results showed a significant interaction between sentence condition (SVO vs. OVS) and target character (Agent vs. Patient) in the adverb region, before the NP2 was encountered. Further analyses confirmed that this was due to the participants looking more towards the object/Patient in the SVO compared to the OVS condition and more towards the subject/Agent in the OVS compared to the SVO condition in this time window.[6]

Altmann and Kamide (2007) use the term *object's affordance* to describe the best possible fit between visual and linguistic information. When presented with a visual scene depicting, for example, a man, a full glass of beer, an empty wine glass, a table, some cheese and party crackers, people anticipated the empty glass upon hearing *The man has drunk*. The empty glass offers the best possible fit for the past tense, although there is nothing to drink inside of it, so it actually does not meet the selectional restrictions of the verb. When, in a second experiment, the stimuli were changed as in (19), participants were more likely to fixate the empty glass in the past tense condition (19-b) and more likely to fixate the full glass in the future tense condition (19-a). Here the difference in fixations as a function of tense became significant for both objects. This could either be the result of the removal of indeterminacy, perhaps the subject in *The man has drunk* had only drunk some of the beer, or because the time window for an anticipatory effect was longer in the second experiment.

(19) a. The man will drink <u>all of</u> the beer.
b. The man has drunk <u>all of</u> the wine.

The authors concluded from their findings that anticipatory eye-movements "reflect an unfolding (mental) world" (p. 515) rather than unfolding language; see also Altmann and Kamide (2009).

Another example of the use of extralinguistic information and object's affordance is provided by a study by Chambers, Tanenhaus, Eberhard, Filip, and Carlson (2002). Here, the semantics of a preposition was used to anticipate a destination. In an act-out task people were instructed, for example, to *Put the cube inside the can*, while they were presented with one or multiple containers (Experiment 1) or with one or two containers that were comparable in size (Experiment 2). The results showed that participants quickly restricted their attention to the only available container, within the 100 ms time window after determiner onset, or, in Experiment 2, the container that could accommodate the cube. However, in the second experiment the proportion of fixations on the target significantly diverged from that on the

[6]Here, the results from the first experiment are reported. Experiment 2 was identical to Experiment 1 only that here distractor objects were added to the scene and sentences were fully cross-spliced.

competitor object only after noun onset but still earlier than in the two compatible referent condition. In Table 4.2, the findings of the studies described above are summarized.

Study	**Source of information**	**Level of prediction**
Knoeferle et al. (2005, Exp.1–2)	visual role information	grammatical/thematic role
Altmann and Kamide (2007)	tense together with visual information	visual and sentence object
Chambers et al. (2002, Exp.1)	lexical-semantic together with visual information	container object
Chambers et al. (2002, Exp.2)	lexical-semantic and pragmatic together with visual information	compatible container object

Table 4.2: Overview of experiments testing the use of additional non-linguistic information for prediction

4.1.3 Prediction based on morphosyntactic information

The following experiments investigated the predictive use of a morphosyntactic cue. A VW experiment developed by Hopp (2012, 2013) could demonstrate that native speakers of German use gender marking on definite articles (20-a) and adjectives (20-b) (Hopp & Lemmerth, 2017) to predict an upcoming noun. The predictive use of gender was reflected by a significant difference in anticipatory looks between same and different gender trials. For the article condition in same gender trials, all identically colored objects had the same gender, whereas in difference trials all identically colored objects had a different gender; for the adjective condition all identically sized objects had either the same or different gender. Thus, only in difference trials was gender informative regarding the prediction of the upcoming noun (for a description of the experimental design, see also section 2.1).

(20) Wo ist . . . (‘Where is . . . ’)

a. der/die/das gelbe NOUN?
the-[*masc.*]/[*fem.*]/[*neut.*] yellow NOUN?

b. ein kleiner/s gelber/s NOUN?
a small-[*masc.*]/[*neut.*] yellow-[*masc.*]/[*neut.*] NOUN?

More recently, Lemmerth and Hopp (2017) employed the same experimental design to test the predictive use of gender with monolingual, simultaneous and successive bilingual children. The German monolingual child group within an age range of seven to nine years was faster in shifting their gaze to the target noun in informative than in uninformative trials, thus demonstrating that they successfully used the gender cue on the article and adjective for prediction. Brouwer, Sprenger, and Unsworth (2017) found that for Dutch, which discriminates between common and neuter gender, children between four and seven years were successful in predicting the gender of the upcoming noun based on the preceding definite article only if they consistently assigned correct gender in production. The researchers split the child group into a non-targetlike subgroup and a targetlike subgroup who had an accuracy of 75% or higher for neuter nouns in an elicited production task. Previous findings had shown that neuter gender is particularly difficult and the grammatical gender distinction in Dutch seems to be acquired relatively late. In their experimental design, also a color term was inserted between article and noun, but only two color-matching objects were displayed, which either had the same or different gender. Like the adult control and targetlike subgroup, who were more likely to look at the target in informative trials before noun onset, the non-targetlike producing children showed a facilitative effect after noun onset.

Hopp (2012) further found that also number marked on articles (21-a) and adjectives (21-b) in German serves as a predictive cue if it is informative regarding the prediction of an upcoming noun. The experimental design was the same as above for gender, where an effect of prediction meant a significant difference between trial types (same vs. difference) before the onset of the target noun. Here number agreement within an NP was investigated.

(21) Wo siehst du ... (‘Where do you see ... ’)

a. den/die roten Eimer?
the-[*sg.*]/[*pl.*] red bucket/s?

b. die kleine/kleinen rote/roten Zahnbürste/n?
the small-[*sg.*]/[*pl.*] red toothbrush/es?

In an eye-tracking experiment with two pictures, Lukyanenko and Fisher (2016) found that number marking on a verb was also used to predict an upcoming noun in English, this way investigating subject-verb number agreement. This finding demonstrates that not only agreement inside an NP is subject to predictive processing. The experiment tested two-and-a-half- and three-year-old toddlers as well as a group of adults. The materials consisted of English *wh*-questions (*Where is/are the [adjective] [noun]?*) or locatives (*There is/are the [adjective] [noun].*), which were compared to sentences that were not informative regarding number marking (*Can you find the [adjective] [noun]?*); see also Table 6.1 in section 6.1. In the experimental condition, participants were presented with a display that showed on one side a large single object and on the other side two smaller identical objects. In the control condition, a display with both objects as singular or plural was shown. Half of the participants were assigned to the experimental and the other half to the control condition, so it was manipulated between subjects. The control condition was introduced to control for differences between the sentence versions and, because previous work had shown that children tend to look more towards the plural picture when the number of both objects differed, to control for a baseline bias. The results of their study showed that three-year-old children and adults had a shorter latency of shifting from distractor to target in informative trials compared to in uninformative trials in the experimental but not the control condition. Also the probability of shifting from distractor to target was higher in informative trials in these two groups. The two-and-a-half-year-old children showed a similar pattern but their latency of shifting to the target, even if they heard the sentences spoken at a slower speaking rate, was only marginally faster and their likelihood of doing so was not significantly higher. The effect seen in the other two groups, however, became visible at noun onset. These findings are in line with those of Kouider, Halberda, Wood, and Carey (2006), who tested the singular-plural distinction in English infants by introducing novel objects, see (22). In an eye-tracking experiment employing the preferential looking paradigm, they found that after a copula followed by a quantifier looks towards the target picture increased for two-year-olds but not for 20-month-old children.

(22) a. Look, there **are some** blicket**s**!

b. Look, there **is a** blicket∅!

When presented with number marking on the noun only (*Look at the blickets!*), three-year-old children but not the two-year-olds looked more towards the target picture, indicating that before the age of three children need multiple cues (verb and noun morphology, perhaps also a quantifier) to succeed.

Another morphosyntactic cue that can be used predictively is case. Kamide, Altmann, and Haywood (2003) could demonstrate that case marking on the first two noun phrases in the head-final language Japanese was used to anticipate the most plausible third NP, the Theme argument. When presented with a visual scene displaying, for example, a waitress, a customer, a hamburger on a table and a dustbin, people directed anticipatory looks towards the hamburger upon hearing a sentence with the sequence 'waitress-*Nom* customer-*Dat*' as in (23-a) but not after 'waitress-*Nom* customer-*Acc*' as in (23-b). No further argument can follow after this last sequence, or at least no plausible Goal argument which could follow was provided in the visual scene. In the dative condition, there were significantly more looks towards the Theme than in the accusative condition. Based on their findings, the authors argue that prediction can be driven by case-structure and this prior to the grammatical head. They also note, however, that in the materials tested lexical-semantic information played a role, and one should probably add world knowledge as well (Kamide, 2008, p. 658ff.).

(23) a. weitoresu-ga kyaku-ni tanosigeni hanbaagaa-o hakobu.
waitress-*Nom* customer-*Dat* merrily hamburger-*Acc* bring
'The waitress will merrily bring the hamburger to the customer.'

b. weitoresu-ga kyaku-o tanosigeni karakau.
waitress-*Nom* customer-*Acc* merrily tease
'The waitress will merrily tease the customer.'

These findings could be replicated by Mitsugi and MacWhinney (2016), who added a scrambled condition with the sequence NP-*Dat* NP-*Nom*. Also in the scrambled condition, anticipation of the Theme argument was observed (for more details see subsection 4.2.4).

To investigate when and how morphosyntactic information is integrated with verb semantics, Kamide, Scheepers, and Altmann (2003) conducted two VW experiments, one in German and one in English. In the German experiment, the NP1 was either the subject or Agent of the sentence and, therefore, nominative marked (24-a) or it was the direct object or Patient and, therefore, accusative marked (24-b).

(24) a. Der Hase frisst gleich den Kohl.
[the hare]-*Nom* eats shortly [the cabbage]-*Acc*
'The hare shortly eats the cabbage.'

b. Den Hasen frisst gleich der Fuchs.
[the hare]-*Acc* eats shortly [the fox]-*Nom*
'The fox shortly eats the rabbit.'

The visual display was the same for both sentence conditions. An interaction between case and object became significant in the adverb region, presumably indicating that there were more anticipatory looks towards the *fox* in the accusative condition and more looks towards the *cabbage* compared to the inappropriate objects in the nominative condition. However, planned comparisons revealed that the difference in looks towards the *cabbage*-objects between the nominative and accusative condition was statistically not significant. In the English experiment, active (25-a) and passive (25-b) sentences were used. Here, unlike in German, the thematic role of the NP1 only becomes evident at the verb.

(25) a. The hare will eat the cabbage.

b. The hare will be eaten by the fox.

The results showed a significant interaction between voice (active vs. passive) and object in the verb region, but also here the comparison between conditions was not significant for the *cabbage*-objects. The authors state that the difference in timing between the two experiments might be the result of integrating two different types of information to form a prediction in German, namely, morphosyntactic information on the NP1 and semantic information on the verb. Hopp (2015) conducted the German experiment again but changed the visual properties of the scenes, which possibly had caused people to look more towards the *cabbage*-objects overall in the previous study. The interaction in the adverb region could be replicated, but now the comparison between looks to the Agent and the Patient was significantly different in both sentence conditions; see also Henry et al. (2017).

Zhang and Knoeferle (2012) investigated how German speaking children at the age of four to five years use case marking to predict an upcoming thematic role when supported by the visual context. They presented the participants with SVO and OVS sentences like (26), while the event described (bull pushes bear) was either depicted or not.

(26)

Den Bär	schubst	sogleich	der	Stier.
[the bear]-*Acc*	pushes	immediately	[the bull]-*Nom*	

For correctly answered trials, children demonstrated anticipation of the Patient in SVO sentences and anticipation of the Agent in OVS sentences in the adverb region when the event was depicted. Children's response accuracy for the comprehension questions showed an effect of case marking and event depiction, although no interaction: The children performed better in the SVO condition and better in the OVS condition when the event was depicted. The researchers further found an interaction between children's accuracy, case marking and event depiction and a trend towards an interaction between working memory score, case marking and event depiction (response accuracy and working memory score were correlated). In the adult group, who did the same experiment, the visual context effect already emerged in the verb region. For them, only case marking had an effect on response accuracy. Thus, when supported by the visual context, the children could override an initial preference for SVO, but were slightly delayed and their online performance was modulated by accuracy and cognitive capacity.

Özge et al. (2016) report in their paper that four-year-old Turkish-speaking children were able to use case marking information, nominative and accusative marked on the NP1, to anticipate the upcoming noun after the verb in verb-medial structures (SVO and OVS) and prior to the verb in verb-final structures (SOV and OSV). The eye-tracking study was modeled after Kamide, Scheepers, and Altmann (2003) described above. Özge et al. used a very similar design to also test German-speaking children. German has a less flexible word order and a less reliable case system than Turkish. Below an example is given of the SOV structures (27-a) and OSV structures (27-b) that were presented to a group of four- to five-year-old German-speaking children (examples and translations follow Özge et al., 2016, p. 295). Comprehension was also tested by a picture verification task following each item.

(27) a.

Der Hase	wird	im	nächsten	Moment	den Kohl	aufspüren.
[the rabbit]-*Nom*	will	in the	next	moment	[the cabbage]-*Acc*	hunt out

'The rabbit will shortly hunt out the cabbage.'

b.

Den Hase	[sic!]	wird	im	nächsten	Moment	der Fuchs	aufspüren.
[the rabbit]-*Acc*		will	in the	next	moment	[the fox]-*Nom*	hunt out

'The fox will shortly hunt out the rabbit.'

The authors analyzed the Agent preference, a categorical variable, defined as the number of samples where participants looked at the plausible Patient minus the number of samples where participants looked at the plausible Agent, depending on condition. In the adverb region, the Agent preference was greater in the accusative (OSV) than in the nominative (SOV) condition, indicating correct anticipation.[7] The authors concluded that German-speaking children do not wait for the verb or simply rely on word order but use morphosyntactic information to incrementally process a sentence and predict upcoming information.

[7] No further information about the picture verification task is given; Özge et al. (2016) only mention a very good performance.

Moreover, they also integrate other sources of information like the visual context and world knowledge. Özge et al. speculate that the reason why their results differ from previous studies reporting difficulties with non-canonical structures might be the experimental design, where a visual context was presented and where the verb-final structures gave the children more time to process. Therefore, the effects found in neurophysiological studies (e.g., Schipke, Knoll, Friederici, & Oberecker, 2012) might reflect the detection of errors or re-analysis and, thus, may be a consequence of children's inability to revise initial parsing expectations.

Some of the studies above demonstrate that certain thematic roles can be predicted, which will be further discussed in the next subsection. These studies appear in Table 4.4; all other studies described in this section are summarized in Table 4.3 below.

Study	**Source of information**	**Level of prediction**
Hopp (2012, 2013); Hopp and Lemmerth (2017); Lemmerth and Hopp (2017); Brouwer et al. (2017)	gender marked on article or adjective	noun
Hopp (2012)	number marked on article/adjective	noun
Lukyanenko and Fisher (2016)	number marked on verb	subject
Kamide, Altmann, and Haywood (2003, Exp.3); Mitsugi and MacWhinney (2016)	case marking	grammatical/thematic role

Table 4.3: Overview of experiments testing the use of morphosyntactic marking for prediction

4.1.4 Prediction of thematic roles

The results of the experiments by Kamide, Altmann, and Haywood (2003, Exp.1), Hopp (2015) and Özge et al. (2016) described above showed that case marking information can be integrated with verb semantics to anticipate a certain thematic role, here a Patient or an Agent. Moreover, the second experiment by Kamide, Scheepers, and Altmann (2003) showed that also thematic role information available on the verb through the marking of voice together with its semantics is used for prediction; for another study on the use of verb semantics together with verb morphology, testing the verb-initial language Tagalog, see Sauppe (2016). Language users moreover combine extralinguistic and case marking information for the prediction of an upcoming thematic role (Zhang & Knoeferle, 2012).

In another eye-tracking experiment, Knoeferle et al. (2005) made use of thematic role ambiguity in German to investigate the effect of a biasing adverb on the prediction of an upcoming thematic role. Here, no stereotypical actions were described (see also Arai & Keller, 2013), ruling out an effect of lexical-semantics and/or world knowledge. In the sentences in (28) the auxiliary *wird* is ambiguous between future tense *will* as in (28-a) and present tense *is* in a passive sentence like (28-b). The adverbs used biased towards either a future tense/active reading or a present tense/passive reading, which was assessed through a pre-test. In the scene presented to the participants, they would see for the example below a princess in the center facing the pirate to her left and with a fencer to her right facing her.[8] In the extended post-adverb region, the time window from 200 ms before the offset of the adverb up to the noun of the NP2 until into the NP2, the interaction between target character (Agent vs. Patient) and sentence condition became significant, with more looks to the Patient for the active than for passive sentences but no effect the other way around. Since it takes approximately 200 ms to launch a saccade, which was not taken into account beforehand, it can be argued that this finding was not due to the preposition/case-marked article of the NP2. Later eye-movements during the second argument showed a clear disambiguation pattern.

(28) a. Die Prinzessin wird sogleich den Pirat waschen.
the princess$_{\text{AGENT/PATIENT}}$ will soon the pirate$_{\text{PATIENT}}$ wash
'The princess will soon wash the pirate.'

b. Die Prinzessin wird soeben von dem Fechter gemalt.
the princess$_{\text{AGENT/PATIENT}}$ is currently by the fencer$_{\text{AGENT}}$ painted
'The princess is currently painted by the fencer.'

As mentioned above, the language user can also employ a combination of cues. Kamide, Altmann, and Haywood (2003) further demonstrated in a VW experiment that English speakers use information about the Agent and the verb to anticipate the most plausible Theme. To control for influences of lexical or conceptual associations, the materials were designed in such a way that effects of combinatory information (information about the Agent and information about the verb) could be singled out. When they listened to the sentence in (29-a), participants looked towards the motorbike more than when listening to (29-c) or (29-b).

(29) a. The man will ride the motorbike.
b. The girl will ride the carousel.
c. The man will taste the beer.
d. The girl will taste the sweets.

Borovsky et al. (2012), moreover, showed that adults and children between the age of three and ten years combined information about the Agent and action to anticipate the most plausible Theme object, for example, *the treasure* upon hearing *The pirate hides*, while also being presented with an agent-related (ship) and an action-related object (bone). Both groups temporarily considered the local competitor. The

[8]The scenes in this experiment were the same as in Experiments 1–2 described in subsection 4.1.2. In Experiment 3, the presented action only became relevant at the end of the sentence after disambiguation.

results further showed that children's predictions were slower than the adults'.

In a follow-up experiment to Altmann and Kamide (1999), Kamide, Altmann, and Haywood (2003) used sentence contexts including ditransitive verbs to investigate the anticipation of the second post-verbal argument. They compared two sentence contexts where a plausible Goal would be either inanimate, as in (30-a), or animate, as in (30-b), while participants saw the same visual scene depicting a woman, a man, butter, bread and a cup as a distractor.

(30) a. The woman will spread the butter on the bread.
b. The woman will slide the butter to the man.

During the first post-verbal argument, the Theme *the butter*, participants initiated anticipatory eye-movements towards the appropriate Goal argument. In addition to a difference between appropriate and inappropriate objects, a further analysis showed a significant interaction between object and verb: Anticipatory looks depending on the verb were confined to animate objects; there was no significant difference for inanimate objects. The authors explain this finding with the fact that the inanimate object *bread* could also be the Theme object of *slide*, which also applied to all other sentence contexts, resulting in an overall high proportion of looks. Unclear is whether participants indeed predicted the upcoming second post-verbal argument at this point or mistakenly expected the Goal argument before the Theme, since some verbs allowed an alternating order (*The man will show the poster to the woman* vs. *The man will show the woman the poster*). Despite this drawback, the results indicated that language users were able to predict a Goal argument when encountering a ditransitive verb.

In another eye-tracking study including three experiments, Boland (2005) investigated the role of situation-based world knowledge and linguistic information. In a first experiment, a main effect of verb argument structure was found. In the dative/Recipient condition (31-a), participants anticipated the Recipient argument irrespective of its typicality, i.e., whether it was more or less typical for it to appear in the event described. Other conditions contained adjuncts that either could be clearly defined as such as in the intransitive/locative condition (31-b) or represented a borderline case in the distinction between arguments and adjuncts as in the action/instrument condition (31-c).

(31) a. The newspaper was difficult to read but the mother suggested it anyway to her teenager/toddler.
(Recipient condition)
b. The girl slept for a while on the bed/bus.
(intransitive/locative condition)
c. The donkey would not move, so the farmer beat it vigorously with a stick/hat every day.
(action/instrument condition)

In a second experiment, when participants were presented with both the typical and the atypical Recipient, a typicality effect emerged.[9] Boland argued that argument structure was used for the identification of potential Recipient arguments and world knowledge served the purpose of selecting the most likely candidate. In both experiments participants also anticipated the typical adjuncts, however, unlike for arguments, Boland assumed that this was the result of co-occurrence patterns and, thus, more local priming (see section 2.2). In a third experiment, she further investigated the special status of argument structure. In sentences like (32), she found more anticipatory looks to the Recipient/Benefactor picture after dative verbs compared to action verbs. Only in the dative verb condition (32-a) is an argument subcategorized.

(32) One window was broken, so the handyman...

a. mentioned it right away to the owners. (Recipient argument)

b. fixed it hurriedly for the owners. (Benefactor adjunct)

pictures: broken window, handyman, owners, tools (potential instrument of the action verb)

An overview of studies demonstrating the prediction of an upcoming thematic role or a feature of a particular thematic role expected to follow is given in Table 4.4.

Study	**Source of information**	**Level of prediction (thematic role)**
Kamide, Scheepers, and Altmann (2003, Exp.1); Hopp (2015); Özge et al. (2016); Henry et al. (2017)	case together with verb semantics	Agent or Patient
Kamide, Scheepers, and Altmann (2003, Exp.2)	voice together with verb semantics	Agent or Patient
Zhang and Knoeferle (2012)	case together with visual context	Agent or Patient
Knoeferle et al. (2005, Exp.3)	semantics of an adverb	Agent or Patient
Kamide, Altmann, and Haywood (2003, Exp.2); Borovsky et al. (2012)	lexical semantics of Agent/subject and verb	Theme
Kamide, Altmann, and Haywood (2003, Exp.1)	verb selectional restrictions	semantic feature of the Goal
Boland (2005)	verb argument structure	Recipient

Table 4.4: Overview of experiments testing the prediction of upcoming thematic roles

[9] Another difference in this experiment was that the stimuli were spoken by another speaker at a faster speaking rate.

4.1.5 Prediction of syntactic structure

In two VW experiments, Arai and Keller (2013) further demonstrated that lexically specific syntactic information of verbs is used to predict upcoming syntactic structure. A first experiment showed that the probability of launching a saccadic eye-movement towards a respective target was higher for transitive (33-a) than for intransitive verbs (33-b)/(33-c) within the duration of the verb, indicating that participants made use of verb argument structure to predict an upcoming direct object (or not).

(33) a. Surprisingly, the nun punished the artist.
b. Surprisingly, the nun disagreed with the artist.
c. Surprisingly, the nun disagreed and the artist threw the kettle.

The visual scene for the example above displayed a nun, an artist and a kettle. In (33-a), participants anticipated the animate entity *artist* upon encountering the transitive verb. Upon encountering the preposition in (33-b), participants also anticipated the *artist*, whereas there was no anticipation in (33-c) for the conjunction. This experiment was intended to provide an extension of the study by Boland (2005) summarized above, but tried to account for some potential flaws in her design. Most notably, in the experiment by Arai and Keller (2013) looks to the same picture were compared and, in addition, verbs were used that are usually not associated with the subject/Agent.

In a second experiment, the authors further found that participants use verb frequency information to predict a relative clause structure. When presented with a verb that infrequently appears in the past participle form like *watch* followed by a reduced relative clause, i.e., without a relative pronoun, like in (34-a), they looked less towards the relative clause Agent (student) than in the unreduced condition (34-b). Presumably, participants first adopted a main clause analysis even if that was implausible (*The videotape watched*) due to the structural ambiguity, showing a garden-path effect. No such difference between the reduced and unreduced condition was observed for verbs that frequently appear in the past participle form (e.g., *record*), indicating that participants anticipated that a by-agent phrase would follow.

(34) a. The videotape watched by the student was found under the chair.
b. The videotape that was watched by the student was found under the chair.

The next experiment described tested the use of syntactic information in children. In a review of the acquisition of syntactic knowledge in children and how this can be linked to parser development, Omaki and Lidz (2015) state that experimental evidence suggests that mechanisms for incremental processing do not qualitatively change during a child's development: They have been shown to incrementally resolve structural ambiguities and actively complete filler-gap dependencies (e.g., Omaki, Davidson White, Goro, Lidz, & Phillips, 2014). However, Atkinson, Wagers, Lidz, Phillips, and Omaki (2018) found that in sentences where prediction was based on successful comprehension of a *wh*-dependency (35-a), five- to six-year-old English children did not perform like adults. In example (35-a) from the materials used by Atkinson et al., there is a direct object (DO) gap after the verb, after which a Patient (cake) should be activated and, presumably, predicted.[10] After encountering the NP, a PP should be activated and generate an expectation for an instrument due to a prepositional object (PO) gap.

(35) a. Can you tell me what Emily was eating the cake with __?
b. Can you tell me if Emily was eating the cake with the fork?

After verb onset, the children showed no effect of question type and fewer target fixations in sentences like (35-a), i.e., looks to *cake*, than adults. For the adults, an effect of question type emerged over time during the verb region. However, children's target fixations increased in the following NP region. Both groups showed an increase in looks to the target instrument, the *fork*, during the NP region for *wh*-questions, but with a steeper slope for the adults. Analyses taking into account age showed that adultlike active dependency formation emerged around the age of six years, however the effect of question type showed up as an interaction with linear and non-linear time in the NP region, that is, still delayed. Vocabulary size modulated the active PO gap analysis, with the high vocabulary group anticipating the target instrument in the NP region. To actively fill a gap, the parser has to keep the fronted constituent, the filler (*wha*t), in working memory and retrieve it at the correct thematic position, here at the DO gap. The authors rule out reduced experience with questions including DO gaps as an explanation for children's performance in DO gap analyses because corpus analyses show that both adult and child corpora exhibit similar distributions of DO gaps. Distributions were also similar for PO gaps. A possible explanation for the findings is that younger children were more conservative in their active dependency formation because of limited cognitive resources that prevent them from 'risky interpretative commitments.' Thus, they may have a higher threshold as to when distributional regularities function as a reliable information source for DO gap analyses. On the other hand, once the object NP is encountered, the PO gap analysis becomes the only viable option, potentially explaining why age could not account for variation here.

In a reply to Phillips and Ehrenhofer (2015), see subsection 3.1.3, Gabriele, Fiorentino, and Johnson (2015) focus on the processing of *wh*-dependencies. They propose that difficulties in active gap-filling may relate to deficits in attentional control, which might account for children's performance. They report a larger subject filled-gap effect, see the example in (36) with a gap after the preposition, for those participants who showed less interference in a number Stroop task and no difference between an adult L1 and adult L2 group in self-paced reading, as further discussed in subsection 4.2.5 for adult L2 speakers.

[10] Note that, at the same time, this sentence is a garden-path sentence. After the verb, the question is complete, however, the Patient of the action is not asked for but the instrument, which only becomes evident when it is overtly realized.

(36) My father asked who Peter will see the boss with __ at the bar.

Further research is needed to back up a relationship between active gap-filling, prediction and cognitive resources. However, it would be in line with the assumption that differences between children and adults can be explained in terms of cognitive immaturity and, thus, reflect the immature parser (Omaki & Lidz, 2015), as discussed in the previous chapter.

In subsection 4.2.5, an ERP experiment by Kaan, Kirkham, and Wijnen (2016) on the use of sentence contexts that allow for an elliptical structure will be described. The experiment showed that, based on this syntactic information, L1 speakers of English can anticipate upcoming syntactic structure, here an ellipsis after a second possessive (e.g., *Although Peter met John's surgeon, he did not meet Max's [. . .]*). An overview of studies that have tested the prediction of syntactic structure (number of arguments/objects, reduced relative clause vs. main clause, ellipsis, gaps) is given in Table 4.5.

Study	**Source of information**	**Level of prediction**
Arai and Keller (2013, Exp.1)	verb argument structure	semantic/syntactic structure
Arai and Keller (2013, Exp.2)	verb frequency	syntactic structure
Kaan et al. (2016)	syntactic information	syntactic structure
Atkinson et al. (2018)	*wh*-dependency	direct object gap

Table 4.5: Overview of experiments testing the prediction of upcoming syntactic structure

4.1.6 Prediction at the discourse-level

Finally, there are cues influencing prediction at the discourse-level, although testing for effects of prediction at this level is quite difficult. One such cue that can influence the prediction of information structure is word order. In the articleless flexible word order language Finnish, the non-canonical order OVS is usually used in those cases where the object is already known but the subject is a discourse-new referent. Kaiser and Trueswell (2004) demonstrated in a VW experiment that, when the critical sentence started with an object, participants anticipated the discourse-new referent, i.e., the character displayed but not yet mentioned.

Language users, moreover, seem to be able to predict the structure of the discourse or coherence. Kuperberg, Paczynski, and Ditman (2011) found a larger N400 after the critical word in causally unrelated scenarios (37-c) than after intermediately (37-b) and highly causally related (37-a) scenarios, irrespective of its position (mid or final); in example (37) the critical word *sunburn* appears mid-sentence. The lexical-semantic relationship between individual words was matched. The third sentence was presented word-by-word. After each three-sentence discourse, participants were asked to judge how easy or difficult it was to connect the last sentence to the previous two.

(37) a. Jill had very fair skin. She forgot to put sunscreen on.
b. Jill had very fair skin. She usually remembered to wear sunscreen.
c. Jill's skin always tanned well. She always put on sunscreen.
She had sunburn on Monday.

The authors take their findings as evidence that, at least when explicitly asked to judge the causal relationship between sentences, build-up of causal coherence at the situation-level (as lexical-semantic co-occurrence was controlled for) immediately influences semantic processing of incoming words. Thus, participants probably use their world knowledge to establish coherence and the results may show predictive inference. However, the experimental design does not rule out other interpretations.

Another cue that can influence discourse expectations is the semantic structure of verbs. A bias towards a particular thematic role has been found for verbs denoting a transfer of possession. Experiments using a sentence or story completion task (e.g., Arnold, 2001; Stevenson, Crawley, & Kleinman, 1994) have shown that people tend to refer to the Goal argument and, thus, to the consequence or end state of the event described in sentences like (38-a) and (38-b).

(38) a. **Jack**$_{\text{GOAL}}$ seized a comic from Bill$_{\text{SOURCE}}$. **He** …
b. Jack$_{\text{SOURCE}}$ handed a comic to **Bill**$_{\text{GOAL}}$. **He** …
c. Jack$_{\text{SOURCE}}$ was handing a comic to Bill$_{\text{GOAL}}$.

Kehler, Kertz, Rohde, and Elman (2008) could show that the preference for the end state is more likely the result of the underlying event structure and not a thematic role bias. They manipulated the event structure through the marking of aspect. There were more Source continuations after imperfect contexts like in (38-c) than after perfective contexts. In addition, they found that after perfective contexts participants predominantly used an Occasion relation, i.e., participants probably expected reference not only with the Goal argument, see (39), but this expectation was bound to the expectation of a particular coherence relation.

(39) Jack$_{\text{SOURCE}}$ handed a comic to **Bill**$_{\text{GOAL}}$. **He** started to read it.

A VW experiment by Grüter, Takeda, Rohde, and Schafer (2016) demonstrated that participants preferred the Goal argument before the referent was re-mentioned and that this preference was modulated by grammatical aspect; for a more detailed description see subsection 4.2.6.

In chapter 8, experiments on the phenomenon of implicit causality, another discourse-level cue, will be reviewed, before an experiment of my own on the use of this cue for prediction is presented. As for the event structure bias, the discourse expectation includes upcoming coherence and reference, see (40).

There, the predominant coherence relation is the Explanation relation, if not indicated otherwise, which draws upon semantics and the referent who is the likely causer/Stimulus (Kehler et al., 2008, p. 26). The VW studies cited in Table 4.6 below found an effect of this bias as the result of prediction.

(40) a. **Jack**$_{\text{STIMULUS}}$ frightens Bill$_{\text{EXPERIENCER}}$. **He** ...
b. Jack$_{\text{EXPERIENCER}}$ fears **Bill**$_{\text{STIMULUS}}$. **He** ...

Study	**Source of information**	**Level of prediction**
Kaiser and Trueswell (2004)	word order	information structure
Grüter et al. (2016)	event structure	coherence and reference
Pyykkönen and Järvikivi (2010); Cozijn et al. (2011); Järvikivi, van Gompel, and Hyönä (2017); Contemori and Dussias (2018)	implicit causality	coherence and reference

Table 4.6: Overview of experiments testing prediction at the discourse-level

4.2 Prediction in L2 speakers

In this section, experimental evidence on prediction in adult L2 speakers will be reviewed. The primary focus is on late L2 learners who have started to learn the L2 at school age and, thus, in a different context than children learning their L1. As motivated by L2 processing accounts that assume more substantial L1/L2 differences, L1 and L2 speakers should differ in their use of grammatical information for prediction and/or their use of discourse-level cues. If, however, prediction in L2 processing is generally reduced, L1/L2 differences should show up whenever L1 and L2 speakers' prediction is compared.

4.2.1 Prediction based on lexical-semantic information

Several VW studies have tested whether L2 speakers use the lexical-semantics of verbs to restrict the domain of subsequent reference, using a design similar to that of Altmann and Kamide (1999). One of these is Chambers and Cooke (2009), who tested whether sentence context and speakers' proficiency attenuate L1-L2 competition effects, here whether participants activate their L1 English when processing French. In their experimental design such an effect would be visible through looks to a competitor picture, a near-homophone of the target word after noun onset. Before the target noun was heard, participants generated anticipatory saccades to the target picture in the restrictive condition (41-a), demonstrating that they can use verb semantics to predict an upcoming object. This effect was moreover influenced by proficiency: The more proficient, the more likely participants were to anticipate.

(41) a. Marie va nourrier la poule.
Marie will feed the chicken

b. Marie va décrire la poule.
Marie will describe the chicken

Irrespective of their proficiency, participants temporarily considered the competitor (*pool*) in the non-restrictive condition (41-b).[11]

To address the concern of individual differences between speaker groups, Dijkgraaf, Hartsuiker, and Duyck (2017) tested a group of L2 learners in their L1 Dutch and L2 English and compared them to a group of monolingual English speakers. The bilinguals started to learn English around the age of ten to eleven years. Here, the researchers also analyzed participants' eye-movements towards a target in restrictive sentences (*Mary knits a scarf*), where the verb restricts the selection of upcoming arguments, versus non-restrictive sentences (*Mary loses a scarf*). Participants saw a display showing four pictures. In the time window for an anticipatory effect, the proportion of fixations on the target was significantly higher for the restrictive condition than for the non-restrictive condition. No difference between L1 and L2 processing was found; there was no interaction between condition and language. There was also no interaction between condition and listener type when comparing English L1-Dutch L1 and English L1-English L2, demonstrating that semantic information could be employed by the monolingual group and the bilingual group in both of their languages. However, there was a difference between monolinguals and bilinguals regarding the time when the effect of condition became significant: The bilinguals were slightly delayed compared to the monolinguals independently of the language they were tested in. The last finding was taken as evidence for the weaker link hypothesis, which states that the link between a lexical item's semantics and phonology is weaker in bilinguals because language use is divided between languages (Gollan, Montoya, Cera, & Sandoval, 2008). No mediating effect of proficiency was found, however all participants were highly proficient.

In a very similar study in English, Ito, Corley, and Pickering (2018), again, found no difference between L1 and L2 speakers in their use of lexical-semantic information as a predictive cue. As in Dijkgraaf et al. (2017), participants saw four object pictures while listening to sentences as shown below with only one object being a plausible continuation in the restrictive condition (42-a).

(42) a. The lady will fold the scarf.

b. The lady will find the scarf.

Here, the authors moreover tested for an effect of cognitive load. Interestingly, cognitive load affected both groups similarly: Those participants who additionally had to perform a working memory task

[11] Also not explicitly stated, other measures indicate that the participants learned French at varying ages and the data might include early bilinguals.

showed a delay compared to the participants who only had to click on the mentioned object irrespective of language group. Ito, Corley, and Pickering note, however, that their L2 group (mixed L1 backgrounds) had been exposed to English for more than 12.5 years on average, and their stimuli sentences were very simple and were spoken slowly with pauses.

In a study by van Bergen and Flecken (2017), the authors were interested in the role of linguistic experience, thereby addressing the question whether predictive processing is also affected by cross-linguistic variation in semantics. In a VW experiment, they tested whether advanced L2 learners of Dutch with different language backgrounds can predict the positioning of an object after encountering a placement verb and whether this depends on similarity to their L1. While German, like Dutch, specifies the object position in placement verbs (Dutch: *zetten*, German: *stellen* – put.STAND and Dutch: *leggen*, German: *legen* – put.LIE), see (43-a) and (43-b), it is not specified in English and French, which use only one general placement verb (English: *put*, French: *mettre*). A visual display included one object (e.g., a bottle) in a standing and a lying position and two other objects (e.g., a ball and a cake) in different positions placed on the same surface (a table or a chair), while one of the following sentences was heard:

(43) a. De jongen zette kort geleden een fles op de tafel.
The boy put.STAND recently a bottle on the table.

b. De jongen legde kort geleden een fles op de tafel.
The boy put.LIE recently a bottle on the table.

c. De jongen plaatste kort geleden een fles op de tafel.
The boy put recently a bottle on the table.

One of the objects displayed in the lying position (e.g., ball) could only be described as appearing in this position due to its intrinsic property. A control condition including the Dutch general placement verb *plaatsen* (43-c), which cannot be used as a predictive cue, was added. The authors analyzed fixations on the two lying objects versus the two standing objects for the time window including the verb to test for effects of prediction, and then afterwards, for target objects that can appear in both positions, fixations on the lying or standing variant of the target object to see whether participants correctly understood the meaning of *zetten* and *leggen*.

Results showed a main effect of verb for the native Dutch and German speakers, demonstrating the prediction of perceptual features of objects, and no interaction with language group. However, whereas the native speakers rapidly differentiated between target object and contrast object (same object, different position) in the verb-object integration window, integration was delayed in the German L1-Dutch L2 group. The last finding was taken as an indication of a higher degree of uncertainty in the German L1-Dutch L2 group. No effect of prediction was found in the English/French L1-Dutch L2 group. There was a main effect of verb in the verb-object integration window, indicating correct identification of the target object, however the effect appeared relatively late. Since the analysis of the verb-integration window as well as an additional control experiment showed that English/French L1-Dutch L2 speakers were able to comprehend and produce Dutch placement verbs accordingly, van Bergen and Flecken view the difference between the L2 groups as resulting from a difference in their specific linguistic experience.[12]

Using the experimental design and stimuli of Borovsky et al. (2012), see subsection 4.1.4 for child and adult L1 speakers, Peters, Grüter, and Borovsky (2015) tested a heterogeneous group of bilingual English speakers on their use of information about the Agent and action. The authors split the group into a higher and a lower proficiency subgroup. Both groups anticipated the target after encountering the verb, thus combining information about the Agent and the action (e.g., *The pirate hides the treasure*, agent-related distractor: *ship*, action-related distractor: *bone*, unrelated distractor: *cat*). The tendency of highly proficient speakers to anticipate earlier was statistically not significant. However, the lower proficiency group was more likely to look towards the action-related distractor after verb onset, which might indicate some kind of uncertainty resulting in the consideration of a locally coherent referent. Activation of the less-likely coherent referent was, on the other hand, also found for the adult and child L1 group in Borovsky et al. (2012). The results from this study show that L2 speakers can also combine the lexical-semantics of a noun and a verb to anticipate a plausible object.

To summarize, the results from studies testing the use of lexical-semantic information do not indicate that L2 prediction is generally reduced but rather show an influence of factors like proficiency and language similarity as they were discussed under subsection 3.2.2. If not encoded in the L1, L2 speakers did not predict the positioning of an object (van Bergen & Flecken, 2017). L2 speakers might also be more uncertain than L1 or more proficient L2 speakers. However, in the study by van Bergen and Flecken (2017) a difference in timing emerged in the integration window, i.e., only after prediction.

4.2.2 Prediction at the levels semantics, morphosyntax and phonology based on prior context

The following two studies investigated whether a lexically constraining context was used by L2 speakers to predict a specific word and its phonological form. In an ERP experiment similar to that conducted by DeLong et al. (2005) described in subsection 4.1.1, Martin et al. (2013) found that late learners of English did not show an increased N400 as native speakers did when encountering an unexpected article

[12]The L2 groups were matched on proficiency (upper intermediate level) but the length of residence in the Netherlands was higher for the English/French L1 group. Two early English-Dutch bilinguals and one early German-Dutch bilingual were included.

(44-b) at the end of a sentence. The sentences were presented in written form and the second part was shown word-by-word.

(44) Since it is raining, it is better to go out with ...

a. an umbrella

b. a raincoat

The L1 speakers further showed a negativity effect after encountering the unexpected noun that was followed by an anterior positivity; the L2 speakers, in contrast, only displayed a greater negative component after an unexpected compared to an expected noun, which was significantly smaller and appeared later than in the L1 group. However, the phenomenon investigated here, phonological agreement, does not exist in Spanish, which was the native language of the L2 speakers being tested. The results are further not fully consistent with those from DeLong et al. (2005), as highlighted by Ito, Pickering, and Corley (2018, p. 2): DeLong et al. (2005) report a graded effect correlating with cloze probability, and Martin et al. (2018) an effect of experimental condition (expected vs. unexpected).

Ito, Pickering, and Corley (2018) note that in the paradigm used in those ERP studies, it could not be controlled for when predictive processing started, which could already happen earlier in the sentence (maybe the word *umbrella* was already predicted upon hearing *rain* due to semantic association) or after the immediately preceding article. For this reason, Ito, Pickering, and Corley employed the VW paradigm to test phonological prediction in L2 processing. Participants were English native speakers and Japanese-English speakers with an AoA ranging from five to 15 years and who had a similar self-related proficiency as those in Martin et al. (2013).[13] The experimental sentences, see (45), included a highly predictable word as assessed through a cloze test. The visual scene appeared 1000 ms before a critical word (in bold) was heard and showed four different objects, with one being the critical one. In the target condition the critical object was the target word (chimney [entotu]), in the English competitor condition an object that phonologically overlaps with the onset of the target word (chick), and in the Japanese competitor condition an object that phonologically overlaps with the onset of the Japanese translation of the target word (pencil [enpitu]). The authors moreover added an unrelated baseline condition (spoon).

(45) The child believed that Santa Claus would come into her house down the **chimney** at midnight.

The participants' task was to respond as to whether the sentence mentioned any of the pictures. Both groups predicted the target word as indicated by anticipatory eye-movements towards the target object, but the effect was delayed in the L2 relative to the L1 group. Moreover, the L1 group was more likely to fixate the English competitor object (chick) than the unrelated object (spoon) before the onset of the

[13]In Martin et al. (2013), the participants self-assessed index was 7.6 in the L2 English on a scale from 1 to 10, where 10 meant native speaker level.

target word, whereas such an effect was absent for the L2 speakers. Crucially, the results demonstrate that L2 speakers did not predict phonological information. Moreover, the further absence of an anticipatory effect for the Japanese competitor (pencil) in the L2 group rules out that there was interference from the L1.

Another study tested the prediction of a noun's gender feature. In an ERP experiment, Foucart et al. (2014) had Spanish native speakers, Catalan-Spanish early bilinguals and late learners of Spanish with French as L1 at an intermediate proficiency level read highly constrained sentences that ended with either an expected (46-a) or an unexpected noun (46-b).

(46) a. El pirate tenia el mapa secreto, pero nunca encontró el tesoro que buscaba.
the pirate had the map secret but never found the treasure-[*masc.*] …
'The pirate had the secret map, but he never found the treasure he was looking for.'

b. El pirate tenia el mapa secreto, pero nunca encontró la gruta que buscaba.
the pirate had the map secret but never found the cave-[*fem.*] …
'The pirate had the secret map, but he never found the cave he was looking for.'

For all three groups the researchers found a modulation of the N400 on the article (*la*) and the noun (*gruta*) that was followed by an anterior positivity. They concluded that, at least when the languages are similar to each other like Spanish and French, L2 learners can predict an upcoming word and its features. A study by Molinaro, Giannelli, Caffarra, and Martin (2017) specifically tested whether there was an effect of the language background. To control for effects of proficiency and competing linguistic information that may arise when comparing monolinguals and bilinguals, the authors compared the prediction of a noun's gender feature in Spanish in early Basque L1-Spanish L2 and Spanish L1-Basque L2 bilinguals. Both groups were primarily exposed to one of their languages up to the age of three. In contrast to Spanish, Basque has no gender system; moreover, determiners appear post-nominally as suffixes (*mahai-a* – the table). The nouns used in their materials were either gender-transparent (*la mesa* – the table) or gender-opaque (*la flor* – the flower); the numbers of feminine and masculine inanimate nouns were counterbalanced. The results of the EEG experiment (recording ERPs and measuring the oscillatory activity time-locked to the determiner), which was otherwise similar to the one by Foucart et al. (2014), showed that the Basque L1-Spanish L2 bilinguals relied more on word form properties, i.e., the noun ending, when predicting a gender-transparent word, demonstrating that distributional regularities of the L1 affected prediction in the L2.

Foucart, Ruiz-Tada, and Costa (2016) could replicate the gender-mismatch effect in spoken Spanish. Late French L1-Spanish L2 speakers demonstrated a negativity after articles that mismatched with the gender of the expected noun. The design was adapted from Foucart et al. (2015), and the materials were the same used in Foucart et al. (2014). Here, the noun was completely muted after article offset, before the sentence continued normally until the end. This was done to control for potential effects of the noun or overlapping effects between article and noun, for example through co-articulation. The participants were told that they would listen to phone calls and also the filler sentences included silence at different positions. After participants listened to the sentence, they were presented with a list of words that were either expected (e.g., *tesoro*) or unexpected, although they did not appear, and words from the fillers or new words matched in length and frequency with the others. They were then asked to indicate whether they had heard these words previously. Interestingly, participants more often (falsely) recognized a word as heard if it was expected and even more so after contexts including the expected article, indicating that anticipation processes create a memory trace of a word. Hence, the results support that L2 like L1 speakers (Foucart et al., 2015) do not only rely on integration processes.

The results from the first two studies cited above indicated that L2 speakers were unable to pre-activate the phonological form of an upcoming noun, although the word itself (its concept) could be predicted (Ito, Pickering, & Corley, 2018). Since the L1 group they were compared to showed a prediction effect, L2 speakers' failure in predicting phonological form, at first glance, seems confirmed. One should note, however, that studies have not been able to robustly replicate pre-activation of the phonological form of a predictable word for L1 speakers when using the ERP method (e.g., Ito, Corley, Pickering, Martin, & Nieuwland, 2016; Nieuwland et al., 2018); see Ito, Martin, and Nieuwland (2017) for an ERP experiment with Spanish L1-English L2 speakers measuring brain responses after noun onset. Recently, this has been taken as support for prediction-by-production, a mechanism that resembles the production process insofar as phonological form is predicted only at later stages after semantics and syntax (see section 2.2). This would explain why, differently, an article that mismatches with the gender of an expected noun elicits an ERP response in L1 and L2 speakers alike. At the time the article is encountered, the noun as well as its gender feature are already pre-activated, so the input creates a conflict as the article encountered cannot agree with the noun that is expected; for (46-b) above the feminine article *la* indicates that the more likely masculine noun *tesoro* (treasure) cannot follow. In the experiments reviewed the source of information covered the previous sentence context, hence, the lexical-semantics of individual words together with world knowledge, but also the build-up situation model, and these probably altogether influenced the prediction, here of a specific upcoming word. One may conclude that L2 speakers have no problems in using lexical-semantics and the prior context and that differences between L1 and L2 speakers can be assigned to the level of prediction, maybe as the result of a reduced availability of time and/or resources as proposed by Pickering and Gambi (2018). It remains unclear how other factors like language similarity may affect the level of prediction, if, for example, also English L1 speakers can pre-activate the gender of an expected noun in their L2 Spanish. Research with balanced bilinguals (Molinaro et al., 2017) who speak two typologically different languages points to an influence of the L1 properties also in L2 prediction. To conclude, prediction based on contextual information in L2 speakers seems to be reduced only at the phonological form level but not at the level of semantics and the level of morphosyntax, at least when the L1 and L2 are similar in the way they encode information (e.g., gender).

4.2.3 Prediction based on the lexical and semantic information of classifiers

In a language like English, nouns can vary depending on their countability, with mass nouns being low in countability; compare, for example, *pencil* versus *milk*. To refer to the number of low countability nouns, an additional item needs to be added as in *two glasses of milk* (Gil, 2013). Chinese and Japanese have a grammatical element for this and differentiate between a set of classifiers: Mandarin Chinese has several hundred, with 50–70 frequently used; Japanese has around 150, with around 30 frequently used (see Grüter et al., 2018a; Mitsugi, 2018). Classifiers are morphemes that co-occur with nouns and assign these to a particular class. They can also function as a predictive cue.

In a VW experiment, Grüter et al. (2018a) tested whether L2 learners rely more on semantics associated with classifiers, which are obligatory after numerals and demonstratives in Chinese, than on information about form class.[14] The nouns associated with a certain classifier are typically also associated with certain semantic properties. In example (47-a) from Grüter et al., the classifier *tíao*, indicating a long, slender and flexible object, appears together with the Mandarin Chinese word for *rope*; however, as shown in (47-b), it can also appear together with the word for *dog*. If only class membership influenced prediction, an anticipatory effect should be observed for both. The materials were adapted from Tsang and Chambers (2011), who tested L1 speakers of Cantonese and found that natives relied more strongly on form class.

(47) a. Nǎ yī tíao shì shéngzi? (prototypical)
which one CL is rope
'Which one is a/the rope?'

b. Nǎ yī tíao shì gǒu? (non-prototypical)
which one CL is dog
'Which one is a/the dog?'

The experimental design was as follows: In the G+S+ condition, participants saw the target, a competitor from the same class and an unrelated distractor. For example, for a sentence like (47-b) a picture of a dog would be the target and a rope would be displayed as a competitor, while the picture of a shoe would be displayed as an unrelated distractor. In the G-S+ condition, a competitor from a different class that shares semantic properties with the target class' semantic property (a wristwatch) would be displayed. In the G-S- condition, a further semantically unrelated object would be shown (an apple). First, all participants

[14]This published article includes data from preliminary testing with 19 L2 participants, including four heritage speakers and speakers with a limited knowledge of classifiers. More recently, the researchers presented data including many more L2 speakers (Grüter, Lau, & Ling, 2018b).

filled in a vocabulary test to check their knowledge of the target classifier-noun pairings. The observed variance in the L1 results thereby indicated that some nouns could occur with more than one classifier.

The authors analyzed the time window from classifier onset to noun onset. When all trials except those with no target clicks were included, a group analysis with the G-S+ condition as baseline showed an interaction between condition and group. There was a significant difference in looks to target between the G-S+ and G+S+ conditions in the L1 group but not in the English L1-Chinese L2 group, indicating that L1 speakers experienced competition from a class-consistent (G+) competitor. L2 speakers, on the other hand, experienced more competition from a semantic property (S+) competitor than L1 speakers. When confining the analyses to only those trials where participants obviously knew the correct classifier-noun pairing, the results did not crucially change (but see Grüter et al., 2018b). As reasons why they did not exclude trials with unexpected/incorrect responses in the vocabulary test in the first run, the authors mention (i) the fact that some nouns appear with different classifiers, (ii) the data loss, and, as noted by the first author, (iii) L2 speakers in their L2 environment cannot just exclude those words they do not know, so these data better reflect L2 processing.[15] The conclusion Grüter et al. (2018b) drew from their findings was that differences observed between L1 and L2 processing reflect differences at the level of lexical rather than grammatical knowledge. Although not explicitly assigned to any proficiency group, the L2 speakers included were not highly proficient as indicated by the self-rated proficiency scores. Note that the statistical models described in Grüter et al. (2018a) tested for differences between conditions. For the L2 group, visual inspection of the data showed an increase in looks to target (vs. competitor) before the noun was heard only in the G-S- condition. As expected, the L1 group showed a similar increase in looks to target in the G-S- condition and the G-S+ condition.

In another VW experiment with only two pictures, Mitsugi (2018) examined the use of Japanese numeral classifiers in L1 and L2 speakers. Similar to Chinese, Japanese has a classifier for one-dimensional long and string-like objects, *-hon*, see (48). As in Chinese, nouns associated with a certain classifier are not always semantically prototypical (e.g., *-hon* is also used more metaphorically for phone calls). Differently, classifier use is less frequent in Japanese and there are more rigid boundaries drawn based on animacy. The participants‘ task was to click on the respective target as in (49).

(48) ni-hon-no banana-o tabe-ta
two-CL-GEN bread-ACC eat-PAST
‘(I) ate two bananas’

(49) soredewa san-hon-no kasa-o kurikku-shite-kudasai
Well-then three-CL-GEN umbrella-ACC Click-do-please
‘Well then, please click on the three umbrellas’

[15] Such an analysis might indeed prove useful when testing lexical information.

Similar to previous studies testing the use of gender, Mitsugi (2018) compared participants' gaze patterns between same and difference trials. In difference trials, the classifier was informative regarding the upcoming noun as both objects displayed belonged to a different class. Two classifiers were used: *-hon* and *-mai*, the latter being prototypically used for thin and flat-surfaced objects; these classifiers are frequently used and appear in first- and second-semester Japanese textbooks. Three native speakers checked that the objects consistently appeared with these classifiers. An offline test assessed whether participants had sufficient knowledge of the classifier-noun parings. The critical window started with the offset of the numeral classifier up to the onset of the noun shifted 200 ms forwards; the noun window was also analyzed, and both regions lasted 1200 ms. Only trials where participants clicked on the correct picture were included.

Growth curve analyses showed an effect of condition and effects of linear and quadratic time for the L1 group: L1 speakers were more likely to fixate the target in informative trials, an effect that increased linearly and non-linearly. There were no interactions with group. A separate analysis for the English L1-Japanese L2 speakers showed the same effect of condition, but only a linear increase and no modulation by the self-rated proficiency scores; overall, participants were assigned to intermediate level. Separate analyses for the noun region showed a vanishing effect of condition in the L1 group, indicating that as soon as the noun was encountered, the target was fixated. In contrast, the effect of condition persisted in the L2 group. An interaction between proficiency and the linear term indicated that the more proficient the L2 speakers were, the faster they shifted their attention to the target noun in uninformative trials. Mitsugi interprets the missing effect of quadratic/non-linear time in the L2 group as an indication that the L2 learners' rate of increase in target fixation was gradual, whereas it would rapidly increase in the L1 group and quickly level off. Another finding, which cannot be explained by the experimental manipulation, was that L2 speakers demonstrated fewer target fixations in the classifier-uninformative condition. Based on further inspection of the data, the author suggests that L2 learners' processing involves more uncertainty and that L2 learners' looking is characterized by wandering around. Altogether, the findings show that L2 speakers can also use the information given by the classifier to predict an upcoming noun, although they might use the transparency of the cue, which was not controlled for in this experiment, which used only prototypical classifier-noun parings.

The two experiments described above are particularly interesting as they investigated a cue where the language comprehender could rely on different information sources: form class or semantics. As the results from Grüter et al. (2018a) indicate, L2 speakers may rely more on semantic associations than L1 speakers, who have been shown to primarily rely on form class (see also Tsang & Chambers, 2011). The results from Mitsugi (2018) show that L2 speakers even when not highly proficient can predict an upcoming noun after encountering a classifier, at least when it is prototypical. Interestingly, they can do so although their L1 English does not have classifiers. As for the use of lexical-semantics reported above, the L2 speakers in Mitsugi (2018) appeared less certain, but again only towards the end of the sentence. To conclude, the results from studies testing L2 speakers' use of classifiers do not fully support the assumptions of the RAGE hypothesis but rather indicate that L2 speakers may draw on different sources of information.

4.2.4 Prediction based on morphosyntactic information

Quite a few studies have tested the predictive or facilitative use of gender marking in L2 speakers, Table 4.7 gives an overview. In an eye-tracking experiment by Lew-Williams and Fernald (2007), the authors manipulated the gender cue on the determiner in Spanish, while participants saw two pictures; for an example, see (50). They found that in different gender trials, when the determiner was informative, Spanish-speaking adults and children between 34 and 42 months identified the target more rapidly (see also Lew-Williams, 2017). The same experiment was conducted with a group of adult L2 speakers by Lew-Williams and Fernald (2010) and with advanced to near-native proficient speakers by Grüter et al. (2012). Interestingly, the facilitative effect could not be replicated for English L1-Spanish L2 speakers.

(50) Encuentra la pelota.
Find the-[*fem.*] ball
informative trial: ball-[*fem.*] vs. shoe-[*masc.*]

Additional experiments in Lew-Williams and Fernald (2010) showed that the L2 speakers could use gender information when presented with the same determiner-novel noun pairs from prior teaching trials but not for familiar nouns, or when generalization was required because the novel nouns appeared together with indefinite articles in the teaching trials but with definite articles in the test trials. Dussias, Valdés Kroff, Guzzardo Tamargo, and Gerfen (2013) tested English L1 speakers with different proficiency levels and a group of Italian L1 speakers in their L2 Spanish. While the native control group shifted their looks earlier to the target in different gender trials, only the highly proficient English speakers and the Italian speakers in the feminine but not the masculine condition did so, indicating that proficiency and language similarity might play a role. Italian also marks gender on the determiner and noun, is partially phonologically transparent and has one feminine definite article *la*, the same as Spanish. Note though that these studies analyzed the facilitation due to the determiner in a time window that also included the noun.

Better support for the predictive use of gender in adult L2 speakers is provided by Hopp (2013) and Hopp and Lemmerth (2017). Hopp (2013) tested a group of German L1 speakers and a group of advanced to near-native L2 learners of German with English as L1 in an elicited production and a VW experiment. In German, the gender paradigm is phonologically non-transparent. The L2 results showed a correspondence between gender assignment in production and gender agreement in comprehension.

If gender was assigned consistently in production, the gender cue on the determiner, when informative, could be used to anticipate the upcoming noun.[16] For both groups, Hopp found a correlation between speed of lexical access measured through a control condition and the size of the gender effect. The design and results for the L1 group were summarized under subsection 4.1.3; an example sentence is repeated in (51). Note that an effect here meant a difference between trial types (same vs. difference).

(51) Wo ist der/die/das gelbe NOUN?
Where is the-[*masc.*]/[*fem.*]/[*neut.*] yellow NOUN?

Study	**Language**	**Group**	**Findings**
Lew-Williams and Fernald (2010), Grüter et al. (2012)	Spanish	L1	✓
		L2, L1 English mixed proficiency levels	✗
		L2, L1 English advanced to near-native	✗
Dussias et al. (2013)	Spanish	L1	fem.✓, masc. ✓
		L2, L1 English highly proficient	fem.✓, masc. ✓
		L2, English L1 less proficient	fem. ✗, masc. ✗
		L2, Italian L1 less proficient	fem. ✓, masc. ✗
Hopp (2013)	German	L1	✓
		L2, L1 English advanced to near-native	✓✗ affected by lexical mastery
Hopp and Lemmerth (2017)	German	L1	article ✓, adjective ✓
		L2, L1 Russian advanced	article ✓, adjective ✓
		L2, L1 Russian high-intermediate	article ✓ ✗, adjective ✓ affected by L1-L2 gender congruency

Table 4.7: Overview of eye-tracking during listening studies testing for a predictive or facilitative use of gender in L1 and L2 processing. The last column indicates whether gender was used (✓) or not (✗).

In Hopp and Lemmerth (2017) the same design was used, but this time an adjective condition was added. In this study, the L2 group consisted of Russian L1 speakers with high-intermediate to advanced proficiency in German. Russian also marks gender but only on adjectives and nominal suffixes, unlike

[16]This is parallel to the findings for child L1 learners of Dutch described in subsection 4.1.3 (Brouwer et al., 2017, p. 63).

German, which marks gender on the pre-nominal article and adjectives. Half of the nouns used had congruent gender in German and Russian. The results showed a link between L1-L2 gender congruency and proficiency: When gender was marked on the article, the high-intermediate L2 group anticipated the upcoming noun only if its lexical gender was congruent in Russian. Overall, in addition to language similarity, the predictive use of gender information seems to hinge on overall lexical mastery, which is associated with proficiency.

While grammatical gender is morphosyntactically realized but is a fixed property of a stem and is stored in the lexicon, grammatical number is considered a morphological marker that combines with the stem it modifies. It is a conceptual feature signaling the quantity of the referent (Barber & Carreiras, 2005, p. 137). L2 speakers have been shown to be sensitive to gender and number agreement violations (e.g., Alemán Bañón, Fiorentino, & Gabriele, 2014; Gabriele, Fiorentino, & Alemán Bañón, 2013; Gillon Dowens, Guo, Guo, Barber, & Carreiras, 2011; Gillon Dowens, Vergara, Barber, & Carreiras, 2010), but effects seem to be moderated by proficiency and probably also language similarity and language asymmetries; for effects of the last factor, see also Tokowicz and MacWhinney (2005). An indication that L2 speakers can also use number marking to predict is provided by a study by Hopp (2012), also described under subsection 4.1.3. The English L1-German L2 speakers, who were at an advanced to near-native proficiency level, could use the number cue on the article as well as on the adjective to anticipate the target noun. Any differences within the L2 group could be explained in terms of their variability in correctly assigning gender. To my knowledge, no published study has tested whether L2 speakers also use verb number marking to predict an upcoming subject noun's number feature.

The use of case marking for prediction seems to be consistently problematic for L2 learners. In a VW experiment, Hopp (2015) tested English native speakers at a low-intermediate to advanced proficiency level in their L2 German and compared the results with those obtained from a group of L1 speakers. He presented them with German sentences like (52), while they were shown a visual scene displaying all three nouns (wolf, deer, hunter) and an unrelated distractor (mountain). The materials were adapted from Kamide, Scheepers, and Altmann (2003) with slight changes.

(52) a. Der Wolf tötet gleich den Hirsch.
[the wolf]-*Nom* kills soon [the deer]-*Acc*
'The wolf will soon kill the deer.'

b. Den Wolf tötet gleich der Jäger.
[the wolf]-*Acc* kills soon [the hunter]-*Nom*
'The hunter will soon kill the wolf.'

Hopp (2015) found for the L1 speakers an interaction between thematic role and word order at the adverb segment, which he took as evidence that L1 speakers integrate morphosyntactic and lexical-semantic information for prediction. The L2 speakers, instead, always anticipated the Patient of the SVO condition

(52-a) irrespective of word order, showing that they relied on verb semantics only. Proficiency did not modulate prediction in the L2 group. The use of case marking was investigated by changing the word order. This resulted in a non-canonical sentence structure in the condition where the object was fronted and, thus, topicalized, which might explain the different outcome for the OVS condition in the learner group. As discussed by Jackson (2008), who found difficulties in the comprehension of temporarily ambiguous object-first constructions for L1 as well as intermediate and advanced L2 speakers with English as L1 (see also Hopp, 2006), object-first sentences are relatively infrequent in German, which shows a preference for subject-initial orders, and, in the absence of further discourse information, represent marked structures; see (53) for an example from her materials.

(53) Welche Ingenieurin traf der Chemiker gestern Nachmittag im Café?
[which engineer]-*Nom/Acc* met [the chemist]-*Nom* yesterday afternoon in the cafe
'Which engineer did the chemist meet yesterday afternoon in the cafe?'

In a VW study conducted by Mitsugi and MacWhinney (2016), word order variation was an experimental condition. They presented participants with Japanese sentences like (54), while showing four pictures displaying the animate Agent and Recipient as well as the inanimate Theme and an inanimate distractor object. The study was modeled after the third experiment in Kamide, Altmann, and Haywood (2003) but added a scrambled condition as in (54-b) and also tested intermediate L2 learners of Japanese with English as L1.[17] Kamide, Altmann, and Haywood (2003) found that Japanese L1 speakers anticipated the most plausible Theme object after the sequence NP-*Nom* NP-*Dat*, whereas they looked less towards the Theme object when the sequence was NP-*Nom* NP-*Acc*.

(54) a. gakkou-de majimena gakusei-ga kibishii sensei-ni shizukani tesuto-o watashita.
school-*Loc* serious student-*Nom* strict teacher-*Dat* quietly test-*Acc* handed over

b. gakkou-de kibishii sensei-ni majimena gakusei-ga shizukani tesuto-o watashita.
school-*Loc* strict teacher-*Dat* serious student-*Nom* quietly test-*Acc* handed over
'At the school, the serious student quietly handed over the test to the strict teacher.'

c. gakkou-de majimena gakusei-ga kibishii sensei-o shizukani karakatta.
school-*Loc* serious student-*Nom* strict teacher-*Acc* quietly teased
'At the school, the serious student quietly teased the strict teacher.'

If language users predicted an upcoming argument based on case marking information, they should also do so when arguments are scrambled. This is indeed what Mitsugi and MacWhinney (2016) found for their native speaker group: Over time they looked more towards the Theme object, the *test*, in (54-a) and (54-b) compared to the monotransitive structure in (54-c) during the critical window. The L2 learners, on the other hand, did not predict, neither in (54-a) nor in (54-b), demonstrating no sensitivity

[17] In section 7.1, word order variation including scrambling will be discussed in more detail. Note that topicalization in German, as seen above in the experiment by Hopp (2015), is assumed to have properties that are distinct from those of scrambling.

to the morphosyntactic cue. However, the L2 learners in this study encountered another difficulty: The comprehension questions that followed each sentence were asked in English. Therefore, they had to switch between languages. Research on code switching shows that it can be associated with processing costs (e.g., Chauncey, Grainger, & Holcomb, 2008; Moreno, Federmeier, & Kutas, 2002). This difficulty also applies to the L1 group, but task-induced strategies as mentioned by Kaan (2014) might have a more severe effect on the L2 group; see subsection 3.2.2. The authors reported that this was done to avoid participants developing a strategy based on case marking.

To summarize the findings for prediction based on morphosyntactic information, previous studies have shown mixed results for the predictive use of gender marking, a predictive use of number marking within NPs, and no effect of prediction based on case marking information in L2 speakers. Hence, the RAGE hypothesis can only partly account for the findings. It has to be stressed that both studies on case included L2 speakers whose L1 has no proper case marking system. The L2 speakers in the study by Mitsugi and MacWhinney (2016) might have been further affected by the code switching. The results indicated that grammatical information was used successfully by L1 speakers throughout but not by L2 speakers in line with the assumption that prediction might be selectively limited. It remains to be seen whether L2 speakers who are familiar with case from their L1 are able to use case marking information predictively.

4.2.5 Prediction based on syntactic information

As mentioned in the prior section on children's prediction, L2 learners have been shown to actively complete filler-gap dependencies. In studies that have found differences between L1 and L2 speakers, these have been in timing (Boxell & Felser, 2017; Felser, Cunnings, Batterham, & Clahsen, 2012). Johnson, Fiorentino and Gabriele (2016), who tested late learners of English with Korean as L1, found no differences in self-paced reading between the L2 and an L1 group. However, they observed a link between working memory and the processing of licit *wh*-dependencies that differed between groups, which raises the question of whether and how cognitive resources might affect L1 and L2 speakers differently. A link between individuals' syntactic processing and attentional control was also discussed in Gabriele et al. (2015) and Johnson (2015). However, the experimental methods used in previous studies do not allow any conclusions about prediction. A design like that used in Atkinson et al. (2018) may provide better support for or against prediction based on syntactic dependencies in L2 speakers.

A study that tested syntactic prediction in L2 speakers was conducted by Kaan et al. (2016), who further investigated how L1 speakers of English and advanced Dutch L1-English L2 speakers integrate syntactic information across clauses. In an ERP experiment, they presented participants with sentences like (55). The sentences (55-a) and (55-c) but not (55-b) and (55-d) allow an elliptical structure, so the expectation for an overt noun after the possessive *Max's* was weaker in these conditions. Therefore, the following preposition *of* should cause a stronger violation in (55-b) compared to (55-a); a noun can be modified by *of*, even if ungrammatical in both contexts. To control whether effects were the result of processing ellipsis and not ungrammaticality, the authors added the conditions (55-c) and (55-d), where the preposition introduces a temporal modifier that does not require a preceding noun.

(55) a. Ellipsis context-*of* (ungrammatical): Although Peter met John's surgeon, he did not meet Max's *<u>of</u> the operation.
b. Non-ellipsis context-*of* (ungrammatical): Although the surgeon met John, he did not meet Max's *<u>of</u> the operation.
c. Ellipsis context-*temporal* (grammatical): Although Peter met John's surgeon, he did not meet Max's <u>before</u> the operation.
d. Non-ellipsis context-*temporal* (ungrammatical): Although the surgeon met John, he did not meet Max's *<u>before</u> the operation.

Grammaticality judgments did not differ significantly between groups. Regarding ERPs, the authors found a positivity for the non-ellipsis conditions (55-b) and (55-d) in contrast to the ellipsis conditions (55-a) and (55-c), starting 500–700 ms after the onset of the possessive and right at the onset of the preposition, for the L1 group. No such effect was found in the L2 group between (55-a) and (55-b), but the L2 speakers showed a positivity for the non-ellipsis temporal condition (55-d) compared to (55-c), which was interpreted as a spurious effect by Kaan et al. These findings might hint at difficulties in the use of syntactic information to predict upcoming syntactic categories in L2 speakers. However, Dutch does not allow an ellipsis after possessive proper names, so an effect of a lack of language similarity cannot be ruled out. Later integration appeared similar in the two groups, however, unlike the L1 speakers, the L2 speakers showed a late frontal negativity in the ungrammatical conditions, which was interpreted as an indication of L1/L2 differences in repair strategies or resources.

4.2.6 Prediction at the discourse-level

To investigate how certain cues affect the interpretation of a sentence or broader discourse, researchers in the field of psycholinguistics have often looked at coreference establishment (e.g., Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000). Grüter et al. (2014) state that "if resources are spread thin during non-native processing, coreference may stand as one of the best candidates for resource allocation given its importance to understanding the speaker's message" (p. 181). Coreference has been shown to relate to the coherence relation a language user establishes. Following Kehler et al. (2008, p. 28), expectations at the discourse-level are coherence-driven, whereby two types of probabilistic information interact: (i) an expectation of how the discourse will unfold with respect to the coherence relation, and (ii) the likelihood that a certain referent will be re-mentioned conditioned on the occurrence of that coherence relation.

In a story continuation task, Grüter et al. (2014), see also Grüter et al. (2017), tested a group of English L1 speakers and a group of Japanese/Korean L1-English L2 speakers on their use of event structure conveyed by grammatical aspect, perfective (56-a) versus imperfective (56-b), for coreference establishment. They also varied the form of reference by presenting participants with either a pronoun or a free continuation prompt. The L1s of the late learners of English show a similar event structure bias; an additional truth-value judgment task further confirmed that the L2 group was sensitive to grammatical aspect in English.

(56) a. Emily$_{\text{SOURCE}}$ brought a drink to Melissa$_{\text{GOAL}}$. (She) ...
b. Emily$_{\text{SOURCE}}$ was bringing a drink to Melissa$_{\text{GOAL}}$. (She) ...

A group interaction in the story continuation task analysis indicated that the L2 group was less likely to use grammatical aspect than the L1 group, whose coreference choices in the pronoun and free prompt conditions was modulated by the perfective/imperfective manipulation: L1 speakers were less likely to refer to the Source argument in perfective contexts (56-a). A main effect of prompt type in both groups showed that participants were more likely to refer to the subject of the previous clause with the pronoun than with the free prompt. A main effect of group indicated that, overall, L2 speakers were more likely to refer to the Goal argument. Interestingly, both groups showed the expected preference for an ongoing-event driven coherence relation for imperfective contexts after a pronoun prompt. The authors took their findings as evidence of a reduced ability to generate expectations (RAGE) at the discourse-level in L2 processing. Grüter et al. (2017) speculated that the L2 group was delayed in their prediction, so they referred more often to the recent Goal argument and only later built a coherence relation. Note that although they are suggestive, results from a story continuation task do not allow firm conclusions regarding participants' prediction. However, a VW experiment presented by Grüter et al. (2016) that tested the online use of event structure confirmed that only the L1 speakers showed a modulating, although small effect of aspect. Both groups demonstrated an overall Goal preference as expected from previous studies testing transfer-of-possession verbs that became visible before the disambiguating pronoun in a time window including 2500 ms of silence; see (57) for an example from Grüter et al. (2016).

(57) Donald$_{\text{SOURCE}}$ brought/was bringing Melissa$_{\text{GOAL}}$ a fancy drink. He/She obviously liked hosting parties.

The authors concluded that the L2 speakers with varying L1s and proficiency levels could generate expectations based on verb semantics but not grammatical aspect, which may indicate a different weighting of cues as further discussed below.

In a similar study, Cheng and Almor (2017) examined the discourse continuations of L1 and L2 speakers after psychological verbs, Experiencer-object and Experiencer-subject verbs. These verbs are associated with two biases: the implicit causality (IC) bias and the implicit consequentiality bias. These verb biases have been found to influence coreference establishment; a detailed description of the phenomena is given in the introduction to Experiment 4 (section 8.1). In L1 research, IC information has been shown to be used to predict upcoming reference (see also subsection 4.1.6). Here, the authors analyzed L2 speakers' sentence continuations only after a pronoun prompt; for a study using the same experimental design but without pronouns, see Cheng and Almor (2019). A group of adult English L1 speakers and intermediate-advanced to advanced Chinese-speaking late learners of English were tested in two different web-based sentence completion tasks. In the first experiment on the use of IC information, participants were presented with sentence fragments up to the ambiguous pronoun including a causal connective as shown in example (58) below. The researchers additionally manipulated the structure of the Experiencer-subject verbs to control for cross-linguistic differences because Chinese typically expresses them periphrastically (58-c). Whereas participants were expected to refer to the NP1 after the Experiencer-object verb *embarrass* followed by the causal connective *because*, they should be more likely to refer to the NP2 after the Experiencer-subject verb *like*.

(58) a. $\text{Paul}_{\text{EXPERIENCER}}$ liked Alan because he ...
b. Ben embarrassed $\text{James}_{\text{EXPERIENCER}}$ because he ...
c. Ben made $\text{James}_{\text{EXPERIENCER}}$ embarrassed because he ...

The results showed that, for the lexical structure (58-a) and (58-b), both groups used IC information to resolve the pronoun, but there were fewer NP2 continuations for Experiencer-subject verbs (58-a) in the L2 group, indicating a weaker bias. No statistical difference between the groups was found for the Experiencer-object verbs. Regarding the structure, the L2 group referred more often to the NP1 antecedent in periphrastic structures (58-c), while no differences between structures were observed in the L1 group. The second experiment was identical to the first one, but here the connective *so* was used, which changes the coherence relation into a Result relation, which should lead to a reversal of the IC bias. The predicted reversal was found in the L1 group; the L2 group, in contrast, showed an overall NP1-preference that was only stronger for the Experiencer-subject verbs. The authors explain the results in terms of a subject/first-mention bias in the L2 group. The reduced use of the semantic and discourse information in this group might result from difficulties in integrating multiple sources of information, leading them to adopt a subject/first-mention strategy as the default. Cheng and Almor also interpret their findings in terms of the RAGE hypothesis but see them as support for the Interface Hypothesis as well.[18]

[18]The authors further view their results as evidence against the Shallow Structure Hypothesis. Following their line of argumentation, this hypothesis predicts that L2 speakers should perform in a nativelike manner based on the statement that L2

However, considering that they tested participants' offline performance only, further research is needed to test this claim with a better suited method.

More recently, a story continuation task and eye-tracking experiment by Contemori and Dussias (2018) with L1 and L2 speakers of English only showed a subtle L1/L2 difference in the online experiment. Although the group analysis showed no between-group differences, the anticipatory effect that only showed up for NP1-biasing verbs was delayed in the Spanish L1-English L2 group.[19] Participants were presented with the two referent AOIs and a distractor; an example item is shown in (59); the authors manipulated whether the continuation was congruent (59-a) or incongruent (59-b) with the IC bias of the verb. A 400 ms pause was inserted between clauses.

(59) Kevin apologized to Dave in the evening/PAUSE because he was scared and because ...
  a. he had insulted him.
  b. he was insulted.

The authors analyzed the time window starting 200 ms after the onset of the pause for 1500 ms. They attributed the delay in the L2 group to a lower quality of the lexical representations of IC verbs, probably as the result of less exposure to English. Note though that here the L2 group tested had a relatively early onset of L2 acquisition with a mean of six years, so the results may be different for late learners.

The results from the experiments described above point to differences between L1 and L2 speakers' prediction at the discourse-level. When presented with a pronoun prompt, L2 speakers were more likely to continue a sentence with reference to the NP1 (Cheng & Almor, 2017). In contrast to L1 speakers, the L2 speakers might consider information like event structure or implicit causality/consequentiality differently. Until recently these assumptions were based on findings from story continuation tasks, which also require people to produce language. While new studies on this topic continued to be published (e.g., Cheng & Almor, 2019; Kim & Grüter, 2019), Experiment 4 of the current thesis was under way, which tested whether L2 speakers have difficulties in integrating different information sources to predict at the discourse-level. There is some indication that L2 speakers, when integrating different sources of information, weigh cues differently. For online processing, this could be shown by Grüter et al. (2016), where the L2 group relied on verb semantics but not grammatical aspect, which is consistent with the assumption that L2 speakers rely less on grammatical but more on non-grammatical information than L1 speakers.

---

speakers rely more on semantic and discourse information relative to (morpho)syntactic information. However, no comparison between different information sources was conducted. Actually the hypothesis does not rule out difficulties due to the integration of cues, but it does not assume the discourse-level to be the source of difficulty.

[19] The fact that only NP1-biasing verbs led to an anticipatory effect in both groups may point to a problem in the experimental design even though an overall preference for the subject/NP1 is very common in studies on pronoun resolution and, in studies with L1 speakers, has also been reported by Pyykkönen and Järvikivi (2010). In the first experiment in Cozijn et al. (2011), the NP1-preference was probably a result of the experimental set-up, as indicated by the results from a second, modified experiment.

4.3 Summary of previous studies

The overview in section 4.1 focused on experimental studies with L1 speakers. Some of the studies demonstrated that also children, who acquire their L1, can predict upcoming information and do so from very early on. However, children are often slower, and as also reported in many studies testing children, language measures like their productive or comprehensive vocabulary (e.g., Borovsky et al., 2012; Lew-Williams & Fernald, 2007; Mani & Huettig, 2012; Nation et al., 2003), developing literacy (Mani & Huettig, 2014) as well as cognitive capacity (e.g., Nation et al., 2003; Zhang & Knoeferle, 2012) probably altogether influence their prediction. A further study one should mention by Gambi et al. (2018) found that children and adults differed in their prediction at the level of phonology. Here, the authors presented participants with two pictures and analyzed their eye-movements after the determiners *one/two* and *a/an*. For the phonological manipulation, target fixations increased only in the adult group; the temporal dynamics of a three-year-old and four-to-five-year-old child group were different. Although the preference for the phonologically predictable picture was evident overall, that is, throughout the critical time window, it did not increase for the children. However, considering that the two experimental sets in Gambi et al. (2018) obviously differed in the reliability of the cue, the conclusion that children have difficulty in predicting at this level should be taken with caution. Fixating a one-object or two-object picture when encountering *one/two* compared to an object starting with a consonant or vowel after *a/an*, opening up several possibilities for upcoming material, appears more straightforward. The experiment by Atkinson et al. (2018) found that at the level of syntax the children tested did not perform in an adultlike manner. It seems possible though that the reason was children's cognitive immaturity. One may preliminarily assume that children can predict but their predictive abilities depend on their cognitive capacity and develop over time. Adults, at least in their L1, are able to use multiple information sources, also in combination, to predict upcoming information at different levels of representation.

Section 4.2 provided an overview of experimental studies with late L2 learners. Previous studies showed no difficulties in the use of lexical-semantic information, as long as there was no difference between the L1 and L2 (van Bergen & Flecken, 2017). Also the use of prior context seemed to pose no challenge for L2 speakers, although they might not predict at all levels of representation: Whereas they were found to predict an upcoming target noun (Ito, Pickering, & Corley, 2018) and its gender feature (Foucart et al., 2014; Foucart et al., 2016), L2 learners of English did not predict at the level of phonology (Ito, Pickering, & Corley, 2018; Martin et al., 2013). Hence, the prediction of phonological form by L2 speakers appeared to be reduced. Experimental findings further indicated that the use of a morphosyntactic cue like case was reduced in L2 processing. The findings for gender were mixed and probably influenced by several factors, including language similarity and lexical consistency, the latter being a factor that is probably also relevant in L1 acquisition (Brouwer et al., 2017). Evidence for (absent) prediction based on syntactic information in L2 processing is sparse and, for the time being, should be taken with caution. Discourse-level cues like event structure and implicit causality/consequentiality appeared to be problematic for L2 speakers, maybe as the result of a difficulty with the integration of different information sources as further discussed below. Overall, recent findings are more in line with the assumption that prediction in L2 speakers is selectively limited. Table 4.8 gives an overview of the findings.

Lexical-semantics/Semantics		
verb information	Theme argument	✓
verb information	object position	✓ (L1 = L2), ✗ (L1 ≠ L2)
subject/Agent + verb information	Theme argument	✓
numeral classifier	noun category	✓
Discourse		
prior context	semantics of upcoming noun	✓
prior context	gender feature of upcoming noun	✓
prior context	phonological form of upcoming noun	✗
event structure	coherence and reference	✗
implicit causality/consequentiality	coherence and reference	✗ (offline)
Morphosyntax/Syntax		
gender marked on article/adjective	upcoming noun	✓/✗
number marked on article/adjective	upcoming noun	✓
case marking	thematic role	✗
syntactic information	elliptical structure	✗

Table 4.8: Summary of experimental evidence so far for late L2 speakers. The left hand column indicates the source of information, the column in the middle the level of prediction and the right hand column whether nativelike prediction was observed (✓) or not (✗).

The L2 studies further showed that L2 speakers' predictive abilities were affected by cognitive load to a similar extent as those of L1 speakers' for a semantic prediction (Ito, Corley, & Pickering, 2018). The observations made by Peters et al. (2015) for less proficient L2 speakers but also the observation by van Bergen and Flecken (2017) for the integration window (see also Mitsugi, 2018) further indicated that L2 speakers' processing displays more uncertainty. However, there was no indication that this uncertainty specifically affected L2 speakers' prediction; it rather affected their later sentence interpretation. The results in Dijkgraaf et al. (2017) indicated that being bilingual rather than being an L2 speaker influenced prediction based on verb semantics. Note that, if not explicitly tested for, this aspect is difficult to control for as most university students, the typical population tested in experiments, speak more than one language but probably with different proficiencies and to varying extents. Interestingly, several studies have demonstrated that L2 speakers showed a different weighting of cues: Grüter et al. (2018a) reported that the L2 group relied more on the semantic information of classifiers, whereas the L1 group primarily relied on form class to anticipate the upcoming noun. Grüter et al. (2016) suggested that the L2 group used verb semantics but not grammatical aspect to predict at the discourse-level. The findings by Hopp (2015) can also be explained in terms of cue weighting. Here, the L1 group integrated lexical-semantic and morphosyntactic information to predict a plausible Agent or a Patient, whereas the L2 group only relied on verb semantics.

The results of previous studies are summarized in Table 4.9. Not shown in the table are studies testing for the predictive use of prior context. Although interesting when investigating the level of prediction, prior context is less suited for systematically testing for the ability to make use of information sources, which will be done in the experiments to follow. Only sources of information are included that were tested in either adult L1 and child L1 or adult L1 and adult L2 speakers. Thus, experiments that, for example, manipulated voice (Kamide, Scheepers, & Altmann, 2003; Sauppe, 2016) are not listed. Not mentioned in the current review but shown in the table are the mixed findings for number marking in adult L1 speakers. This topic will be addressed in detail in the introduction to Experiment 2 (section 6.1). The indicated age of the children in the table shows the youngest age at which prediction or facilitation emerged in the respective studies, however the age might vary depending on the language under investigation.

Studies	Source of information	Adult L1	Child L1	Adult L2	Comment
Altmann and Kamide (1999); Nation et al. (2003); Chambers and Cooke (2009); Kukona et al. (2011); Mani and Huettig (2012, 2014); Gambi et al. (2016); Dijkgraaf et al. (2017); Ito, Corley, and Pickering (2018); van Bergen and Flecken (2017)	Lexical-semantics of a verb	✓	✓ (age 2)	✓	affected by language similarity
Kamide, Altmann, and Haywood (2003); Borovsky et al. (2012); Peters et al. (2015)	Information about Agent and action	✓	✓ (age 3)	✓	indication of more uncertainty at lower proficiency level
Gambi et al. (2018)	Numeral (*one/two*)	✓	✓ (age 2)	???	can probably be used in L2
Grüter et al. (2018a, 2018b); Mitsugi (2018)	Classifier	✓	???	✓	L2ers rely more on semantic information
Gambi et al. (2018)	Phonological information (*a/an*)	✓	✓ (age 3)	???	adult L1 $\neq$ child L1
Lew-Williams and Fernald (2007, 2010)*; van Heugten and Shi (2009)*; Grüter et al. (2012)*; Dussias et al. (2013)*; Hopp (2012, 2013, 2016); Hopp and Lemmerth (2017); Brouwer et al. (2017); Lew-Williams (2017)*; Lemmerth and Hopp (2017)	Gender	✓	✓* (age 2-3) ✓ (< age 4)	✓/✗	affected by lexical consistency, proficiency and language similarity in L2
Kouider et al. (2006)*; Robertson, Shi, and Melancon (2012)*; Hopp (2012); Riordan et al. (2015); Lukyanenko and Fisher (2016)	Number	✓/✗	✓ (age 3)	✓ (within NP)	not tested across phrases in L2
Kamide, Altmann, and Haywood (2003); Kamide, Scheepers, and Altmann (2003); Zhang and Knoeferle (2012); Hopp (2015); Mitsugi and MacWhinney (2016); Özge et al. (2016)	Case	✓	✓ (age 4-5)	✗	effect of language similarity unclear
Atkinson et al. (2018); Kaan et al. (2016)	Syntactic information	✓	✓ (age 6)	??? (filler-gap) ✗ (ellipsis)	adult L1 $\neq$ child L1 (filler-gap), effect of language similarity unclear
Grüter et al. (2016)	Discourse-level information (event structure)	✓	???	✗	prediction based on verb semantics but not grammatical aspect in L2
Pyykkönen and Järvikivi (2010); Cozijn et al. (2011); Järvikivi et al. (2017)	Discourse-level information (implicit causality)	✓	???	???	indication of reduced effect in L2 in offline tasks
*facilitative effect					

Table 4.9: Overview of experimental findings for adult L1, child L1 and adult L2 speakers grouped according to the source of information. The tick (✓) indicates an effect of prediction/facilitation, the cross (✗) its absence. The indicated age for the children shows the youngest age at which prediction or facilitation emerged.

Chapter 5

Experiment 1: Prediction based on the lexical-semantics of the verb

5.1 Introduction

The classic example often cited to refer to prediction during language comprehension is the VW experiment by Altmann and Kamide (1999). Here, participants demonstrated anticipatory eye-movements towards the post-verbal argument when encountering a verb that restricts the selection to only one object in the visual scene; see condition (a) in Table 5.1. Since then, this experimental design has been adopted by many other researchers. The predictive use of the selectional restrictions a verb poses on an upcoming argument has been shown for children from the age of two years onwards (Mani & Huettig, 2012, 2014) and, as found for adults by Kukona et al. (2011) and for children at the age of three years by Gambi et al. (2016), results from the use of verb argument structure rather than only simple association. Experiments indicate that L2 speakers also use this source of information to predict upcoming linguistic input (Chambers & Cooke, 2009; Dijkgraaf et al., 2017; Ito, Corley, & Pickering, 2018). In the 'classic' experimental design, the restrictive condition is compared to a neutral condition, see condition (b) in Table 5.1, including a verb that is not restrictive. Thus, there is a baseline where looks to the target object should not significantly differ from looks to the other objects displayed until the onset of the post-verbal argument. More target looks in the restrictive condition are then taken as evidence for a predictive use of lexical-semantic information.

	subject	verb	CRITICAL WINDOW	direct object
(a) restrictive	The boy	will eat	the	cake
(b) non-restrictive	The boy	will move	the	cake

Table 5.1: The classic experimental design by Altmann and Kamide (1999). Looks to one object are compared between a restrictive condition and a non-restrictive or neutral condition.

Previous experiments have mixed the selectional restrictions so that, for example, an object that can be eaten, climbed on or arrested has been the target in a restrictive sentence condition. The following experiment differs from previous studies on the use of lexical-semantics of verbs insofar as there was no comparison between a restrictive and non-restrictive or neutral condition, where either one or no particular picture should be anticipated. Instead, the verb carried information about a specific semantic feature of the post-verbal argument or direct object, so either an animate or an inanimate object in the visual display should be preferably fixated within a critical time window for an anticipatory effect. This way, a more systematic approach than in previous VW studies was adopted by focusing on only one selectional restriction a verb can pose on an upcoming argument, similar to the ERP experiment with L1 speakers conducted by Szewczyk and Schriefers (2013) or the first experiment in Kamide, Altmann, and Haywood (2003). Note that this should in no way imply that previous studies have been badly designed. In fact, these studies have had the advantage that looks to the same target picture were compared; however, the level of prediction was not systematically controlled for.

Nouns in language can be roughly categorized as being either animate or inanimate. Typically, animacy is described along a scale from human to animal (animate) to inanimate. This animacy hierarchy is also reflected in corresponding marking in some languages (see, e.g., Comrie, 2006). The distinction can be even more gradient with humans ranking above animals, then plants and so forth (Yamamoto, 2006). Typological and psycholinguistic studies show a general preference for animate entities in the beginning of a sentence, known as the Animate First Principle. Animacy can play a role in the processing of many linguistic phenomena. For instance, it can influence ambiguity resolution (*the defendant/evidence examined by the lawyer*) as shown by Trueswell, Tanenhaus, and Garnsey (1994) and the processing of relative clauses (e.g., Mak, Vonk, & Schriefers, 2006), and it strongly interacts with the assignment of thematic roles or, in Turkish, number agreement (Bamyacı, Häussler, & Kabak, 2014). Because it interacts with many other factors, the influence of animacy sometimes remains unrecognized (see de Swart, Lamers, & Lestrade, 2008). The animate-inanimate distinction seems special compared to other semantic features and has been considered a ‘grammaticalized’ semantic feature. To examine the special status of animacy, Szewczyk and Schriefers (2011) contrasted semantic and animacy violations in an ERP experiment. The results showed a biphasic pattern with no difference in the amplitude of the N400 but a higher amplitude of the P600 for animacy violations, while it was controlled that both violation conditions had zero-cloze probability. The authors explain their findings as resulting from animacy being “deeply enrooted in the organisation of language processing” (p. 2016). Note though that in Polish, the language under investigation in that study, animacy is realized morphosyntactically on the direct object, which was the critical word. Altogether, animacy is an important and universal feature in language processing and, hence, a semantic feature likely to be predicted.

Findings from recent experimental studies have gathered more information on L2 speakers' use of information sources and the levels of prediction. Experiments examining the use of prior discourse context, including the lexical-semantics of single words, to predict the phonological form of an upcoming noun have shown that prediction at this level is difficult for L2 speakers (Ito, Pickering, & Corley, 2018; Martin et al., 2013) and probably is not just a consequence of a lack of language similarity, a factor highlighted by Foucart et al. (2014). Pickering and Gambi (2018) argue that the difficulty in predicting at the form level is the result of a hierarchy flow in which predictions are generated based on prediction-by-production, that is, from semantics to syntax and then to form. Later stages of the production process are thus susceptible to difficulties. This is in line with the assumption that L2 speakers' processing might not be as fast as L1 speakers'. A study by van Bergen and Flecken (2017) further showed that also in the use of lexical-semantic information of verbs, language similarity or experience can play a role. In the following experiment, there were no differences between the L1 and L2 of the non-native group for the source of information: The Russian translation of a German verb also encodes the information that the upcoming noun is highly likely to be either animate or inanimate. Furthermore, the experiment examined prediction at the level of semantics. Since the pictures were shown visually, participants were expected to already pre-activate more specific lexical items. Prediction at the level of semantics should be unproblematic for L2 speakers. Even under cognitive load, no differences from L1 speakers were observed for prediction based on verb semantics in a recent study by Ito, Corley, and Pickering (2018). Nevertheless, another recently published study by Ito, Pickering, and Corley (2018) reported a different timing in L2 speakers and, as discussed in subsection 3.2.2, L2 speakers might be affected by a range of factors that probably slow their processing in the L2.

The first experiment in a series of experiments on prediction in late L2 speakers presented here tested the use of lexical-semantic information to predict the animacy, a semantic feature, of the upcoming direct object noun. The object directly followed the verb without any intervening material like an adverbial, hence there was also no long distance between the lexical-semantic cue given at the verb and the predicted entity. This was done for several reasons: First, to demonstrate that the L2 population tested throughout can, in general, predict. Second, to motivate that, as I hypothesized, the linguistic domain might be decisive regarding whether L2 speakers predict. Based on the prior literature review and the assumption that prediction in L2 speakers is only selectively limited, L2 speakers should show a nativelike pattern for prediction based on non-grammatical information, whereas L1/L2 differences might emerge for prediction based on grammatical information or when information sources from different linguistic domains have to be integrated. Therefore, Experiment 1 should answer the following research question:

> RQ: Do L2 speakers of German use the lexical-semantics of verbs to predict the animacy of the upcoming direct object noun?

Based on the results of previous studies referred to above, the predictions were as follows: For prediction at the level of semantics based on a cue that exists in both the L1 and the L2 it was expected that the two groups would pattern alike. The L1 speakers were expected to anticipate the respective target as soon as the verb was encountered, that is, immediately after verb offset. Hence, a predictive use of the lexical-semantic cue should be reflected in more/increasing looks to the target relative to a competitor picture in the critical window, i.e., the time window between the offset of the verb and the onset of the target noun. If an effect of prediction was absent, however, there should be no difference between the probability of looks to the target and looks to the competitor before the onset of the target noun.

5.2 Methods

Experiment 1 was conducted together with Experiment 2 and Experiment 3 in one experimental session, so the participants and the procedure were the same. The experimental items appeared within the same presentation list.

5.2.1 Participants

The L1 group consisted of 28 German native speakers (23 female, 4 left-handed, 1 ambidextrous), none of whom reported being early bilingual. Four additional participants for the L1 group had been tested but were excluded due to bad calibration and/or unstable eye-tracking. In the L2 group, data from 25 Russian native speakers (22 female, 3 left-handed), none of whom reported being early bilingual, was included. Five additional participants for the L2 group had been tested but were excluded: Two were excluded due to bad calibration and/or unstable eye-tracking. The others were excluded due to their accuracy in the behavioral task of the eye-tracking experiment (below 80%) (one participant), or their German proficiency (score below 21 in the Goethe test) (one participant), and/or their performance in the offline questionnaire (below 80% correct responses) (one participant). All L2 participants included in the analyses had started to learn German at or after the age of seven. They were highly proficient in German as measured by the Goethe placement test (Goethe-Institut, 2011), a 30-item multiple choice test, placing them at upper B2- to C1/C2-level according to the CEFR (Council of Europe, 2001). Their length of residence in Germany and, thus, their years of immersion ranged from zero, for a person who was only visiting Germany but had studied German home at Moscow, to 21 years. For an overview of the participant details, see Table 5.2.

All participants had normal vision or, when necessary, wore glasses or lenses and reported no speech or hearing disorders. They either received course credit or a payment of six euros, for the L1 group, or of eight euros, for the L2 group. All participants gave informed written consent. All procedures were in accordance with the Declaration of Helsinki. The testing took place in Potsdam and Berlin.

L1 group (n = 28)			
	Mean	SD	Range
Age	25.96	5.81	18–48
L2 group (n = 25)			
	Mean	SD	Range
Age	27.64	4.69	19–37
AoA German	13.08	4.97	07–24
Years of exposure to German	14.56	6.98	03–30
Years of immersion	07.92	5.89	00–21
Goethe test	25.32	2.23	21–29

Table 5.2: Participant details for Experiments 1–3, showing the means, standard deviations and ranges

5.2.2 Design & materials

5.2.2.1 Pre-test

First, a cloze test was conducted to evaluate the materials. For the materials, 34 sentence pairs were constructed that included a verb that should bias either towards an animate or an inanimate direct object as shown in (60). All sentence contexts started with a definite NP and a restrictive verb; for the example (60) below it was assumed that *füttern* (feed) restricts the selection to an animate noun and *bügeln* (iron) to an inanimate noun, followed by a definite article and an adjective.

(60) a. Die Frau füttert die schwarze . . .
The woman feeds the black . . .

b. Die Frau bügelt die schwarze . . .
The woman irons the black . . .

The resulting 68 sentence contexts together with ten fillers were presented to 35 people in a randomized order in a cloze test. Here, Google forms were used and the link to the test distributed to a group of people who were naive to the purpose of the experiment. For this fast check no demographic information was assessed from the participants, who were instructed to write down the first word that came to their mind. They were given as an example the sentence in (61) and told that this sentence could be completed with *Kuh* (cow) or *Ziege* (goat). The ten fillers consisted of sentences where participants should have a clear intention and, thus, functioned as control sentences, see (62). As expected, participants for the most part answered the sentences with the exact noun intended or another word that would fit the context.

(61) Der Bauer melkt morgens die gefleckte . . .
The farmer milks in the morning the spotted . . .
'In the morning, the farmer milks the spotted . . . '

(62) Der Student stellt das Buch in das . . .
The student puts the book on the . . .

For the eye-tracking materials, sentence contexts were excluded if a plausible direct object was mentioned that had a different semantic feature than the one expected or if a plausible direct object was mentioned, to which no clear animacy status could be assigned; examples of such problematic contexts are shown in (63).[20] After exclusion of these sentences, 24 sentence pairs were left for Experiment 1.

(63) a. Der Chef schreibt (vs. begrüßt) den wichtigen Geschäftspartnern.
The boss writes (vs. welcomes) the important business partners
expected noun: letter or other kind of document

b. Der Polizist entschärft (vs. verhaftet) die schlimme Situation.
The policeman defuses (vs. arrests) the bad situation
expected noun: bomb

5.2.2.2 Eye-tracking materials

For the 24 sentence pairs selected from the pretest, the animate or inanimate object that had the highest cloze probability was chosen as the direct object of that sentence. Overall, the cloze probability for the direct object noun was 54% (range: 14%–86%) in the animate and 51% (range: 26%–91%) in the inanimate condition.[21] An example of a sentence pair is given in Table 5.3.

	subject/Agent	verb	CRITICAL WINDOW	direct object noun
(a) animate	Die Frau The woman	**füttert** feeds	die schwarze the[fem.] black[fem.]	Katze. cat.
(b) inanimate	Die Frau The woman	**bügelt** irons		Bluse. blouse.

Table 5.3: Auditory stimulus in Experiment 1. Each item appeared in two conditions, including either an animate-biasing (a) or an inanimate-biasing verb (b). The critical window was cross-spliced.

In German, the possible continuation is moreover restricted by the gender marked on the article and adjective, which in the current design was always the same for the animate and inanimate direct objects. The sentences were recorded in a sound attenuated room and spoken by a female native speaker of

[20] An additional problem with the excluded sentence pair in (63-a) was that for the verb *schreiben* (to write) a masculine singular accusative object or plural dative object could follow after the article-adjective segment. This was not the case for the sentence pairs that were finally selected.

[21] It should be mentioned that for some contexts the participants completed the sentence with a word that could be depicted by a picture very similar to or even the same as the one that was finally used. For instance, several participants wrote *pony* and the picture displayed in the eye-tracking experiment showed a horse, the word the majority of participants decided on.

German at a normal speaking rate. The article-adjective segment constituting the critical window was cross-spliced in Praat (Boersma, 2001), so it had exactly the same length and prosody in both conditions.

The picture set consisted of non-colored drawings taken from the MultiPic database (Dunabeitia et al., 2018) complemented by new drawings and/or drawings taken from https://openclipart.org that were open source. All pictures of human beings had the same style throughout the experiment. The pictures were saved as .png with a size of 400 x 400 pixels. The subject and Agent of an experimental sentence was always displayed at the bottom center of the screen. It was counterbalanced whether the target object appeared on the left or right side above the subject/Agent. Figure 5.1 shows the visual display, which was kept the same in both conditions, for the item in Table 5.3.

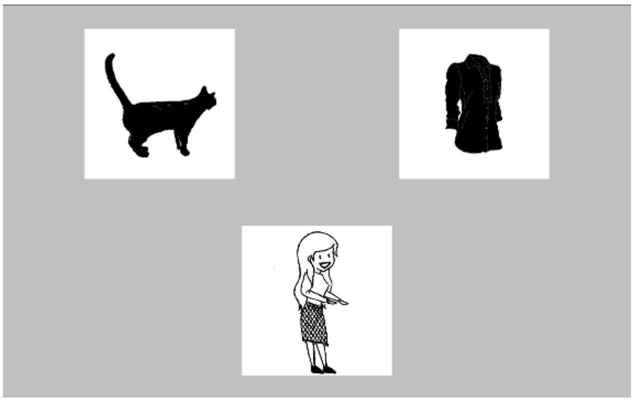


Figure 5.1: Visual display in Experiment 1. The picture at the bottom center shows the subject/Agent and the two upper pictures the target object and the competitor object.

The items were equally distributed across two lists using a Latin square design, so each participant encountered an item only once, either in the animate or inanimate condition. Two further experimental sets (see Experiment 2 and Experiment 3) and additional fillers, 84 additional sentences per list in total, were added. An example of a filler sentence is shown in (64); here the sentence also includes a restrictive verb but is followed by a plural. The order of presentation was pseudorandomized. All experimental sentences and fillers are listed in Appendix B under section B.1 and section B.4.

(64) Der Polizist verhaftet schließlich die Diebe.
The policeman arrests finally the thieves
'Finally, the policeman arrests the thieves.'

With regard to potential differences between Russian and German for the experimental construction, one should note that animacy can be marked through differences in case marking in Russian. For example, animate masculine singular nouns can form the accusative like the genitive (e.g., Bruns, 2007; Cubberley, 2002). Thus, for some sentences the Russian translation would have included a genitive-marked adjective, which is informative regarding the animacy of the upcoming noun. However, no such marking should be expected for German. Moreover, after some verbs a dative instead of an accusative object would follow in the Russian translation. In addition, Kempe (1999) found that languages differ regarding to which extent they use animacy as a cue in sentence processing, reporting such a difference for German and Russian, with animacy being a stronger cue in German to identify an Agent in a simple active transitive sentence and case marking a stronger cue in Russian. Since Experiment 1 addressed the question whether L1 and L2 speakers anticipate an animate or inanimate direct object, while the first-mentioned subject and Agent was always animate, the different weighting of case and animacy information for the identification of an Agent was not expected to play a role in the current design. In addition, the crucial information, namely the lexical-semantics of the verb, does not differ between German and Russian; only the post-verbal argument might differ in case marking. However, this information was encountered after the verb and was not expected to influence the prediction process, which should be initiated earlier and include the semantic feature of the noun. Crucially, the verb denoted the same action in Russian and German.

5.2.3 Apparatus

Eye-movements were tracked with an SMI RED eye-tracker at a sampling rate of 120 Hz. Participants sat at a distance of approximately 65 cm in front of the stimulus screen, which had a resolution of 1680 x 1050 pixels. Both eyes were tracked and their movements averaged. A 9-point calibration procedure was used. The goal was to have a visual acuity below 0.5 degrees, but since the viewing was not restricted and the picture size was quite big, calibration was sometimes accepted when below 0.99 degrees and eye-tracking looked otherwise stable.

5.2.4 Procedure

5.2.4.1 Experimental procedure

The L1 speakers only participated in the visual-world eye-tracking experiment. An experimental session took around 30 minutes. The L2 speakers additionally filled in a questionnaire afterwards that evaluated their general knowledge of the linguistic phenomena tested in Experiment 2 and Experiment 3 and checked the vocabulary used in the experimental items; for Experiment 1 this included a list of the verbs encountered by the participant, who had to indicate whether she was familiar with them.[22] For the L2 speakers, the experimental session lasted around 45 minutes. All participants received the same written instructions (see section B.5 in Appendix B) and each experimental session started with four practice trials, after which participants could ask questions. The participants listened to the spoken sentences via headphones. Each trial started with a preview of 1000 ms. After each sentence the visual display remained on the screen for 800 ms. All participants saw 108 trials, of which 50% were followed by a written statement referring to the prior spoken sentence, requiring a true-false judgment (e.g., *The patient*

[22] I will henceforth refer to all participants as ‘she.’

has been known for a longer time, while the sentence said *new patient*). To respond with either *yes* or *no*, participants had to press a button on a gamepad. They had no time pressure and were encouraged to respond as accurately as possible. The number of *yes* and *no* responses was counterbalanced. Each experimental session contained two breaks, which allowed the participants to rest. Each break was followed by new gaze calibration and validation, which was repeated whenever necessary.

5.2.4.2 Data pre-processing

All trials with a verb that an L2 speaker indicated to be unknown in the vocabulary list were excluded on a by-participant basis (24 trials in total, 4% out of all observations in the L2 data). Moreover, the third block after the second break for one of the L1 speakers (1% of the L1 data) and the first block up to the first break for one of the L2 speakers (further 2% in the L2 data) were excluded, in the first case because eye-tracking became unstable towards the end of testing, and in the second case because the participant still showed difficulties with the experimental task after the practice trials. Participants could fixate either one of three AOIs or the background of the screen. The size of an AOI corresponded to the size of the picture, meaning that looks outside, i.e., the grey part in Figure 5.1, counted as looks to the background. Background looks were treated as trackloss (within the critical window 3% for L1 and 4% for L2). It was expected that participants would fixate the picture displaying the subject/Agent upon its mentioning before moving their eyes towards the critical pictures when encountering the verb. Figure 5.2 shows an overview of the time course including all AOIs per group and animacy-bias condition: The proportion of looks for the animate object (red line) and inanimate object (blue line) clearly diverged before noun onset. At the beginning of a trial, participants fixated the subject/Agent (grey line) as intended.[23]

Since they were not of relevance for the research question, fixations on the subject/Agent within the critical window were discarded and only fixations on the animate and inanimate object, starting 200 ms after verb offset and ending 200 ms after the offset of the adjective and the onset of the direct object noun, were considered for the growth curve analysis. The critical window had an average length of 529 ms (SD: 84.69 ms). Here the exact on- and offset information determined in Praat (Boersma, 2001) was used. Data points were aggregated into 50 ms time bins. In the L1 group, 77% of the data within the critical window included trials with fixations on the animate or inanimate object, and in the L2 group 80%. The amount of discarded data was similarly distributed across conditions. For all analyses and previous data preparation, the eyetrackingR package (Dink & Ferguson, 2015) was used.

[23]There was still a high proportion of looks for the picture showing the subject/Agent during the critical window. Future experiments could insert a filler phrase or, if possible, a natural pause after a gaze neutralizer.

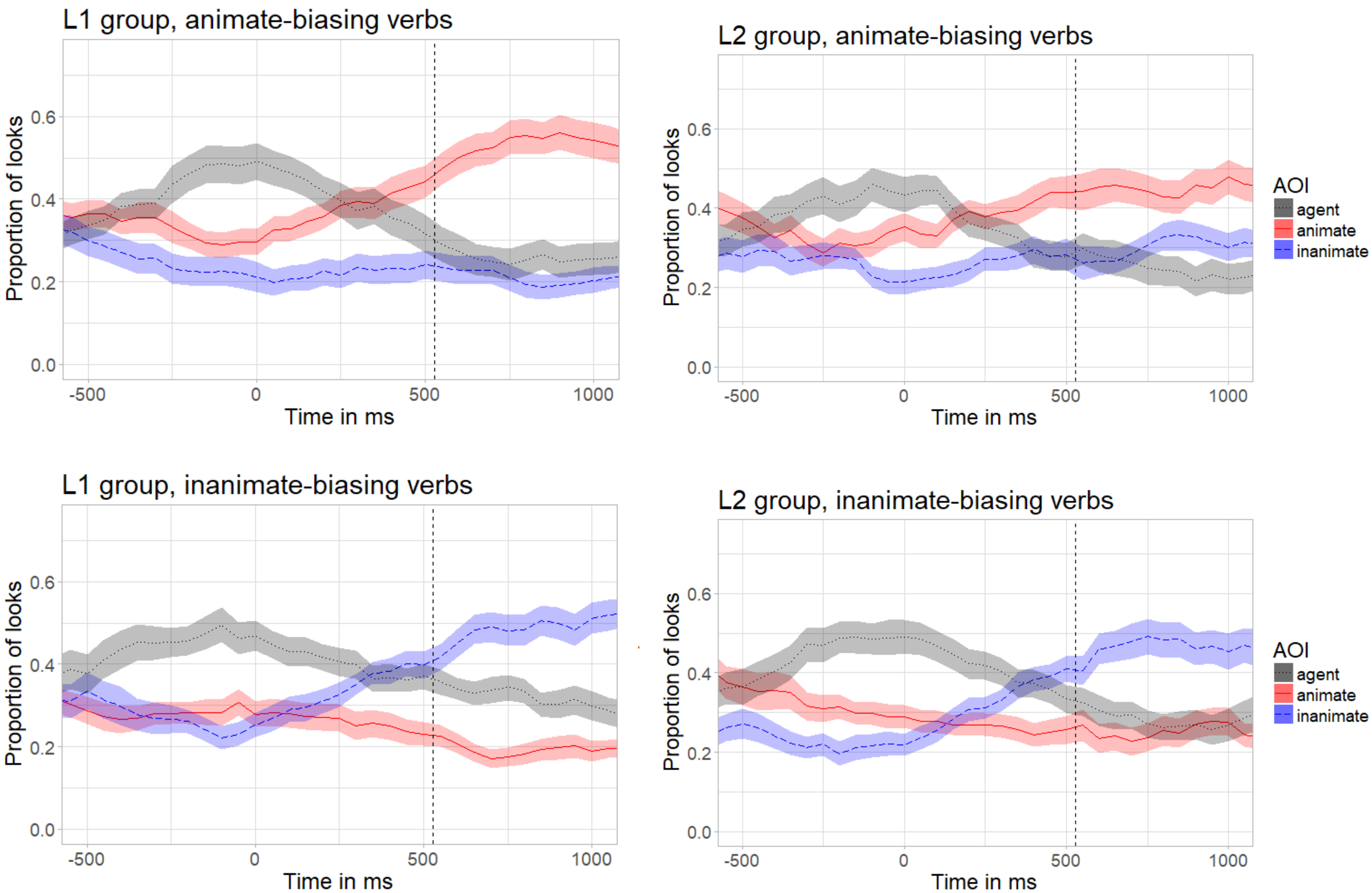


Figure 5.2: Proportion of looks for all AOIs in Experiment 1 for the L1 group (left) and L2 group (right). The plots show the proportions for animate-biasing verbs (top) and inanimate-biasing verbs (bottom); the verb offset shifted 200 ms forwards is aligned to zero, and the dashed vertical line marks the mean onset of the target noun.

5.3 Results

5.3.1 Behavioral data

The L1 group showed an overall response accuracy of 97% in the truth-value judgment task (SD: 4%, range: 80%–100%). The L2 speakers correctly responded to the statements with a mean of 93% (SD: 5%, range: 80%–100%). The results thus indicate that participants paid attention and had no problems in understanding the content of the sentences. Only one participant was excluded on the basis of the accuracy in this task.

5.3.2 Eye-tracking data

In the analysis of the eye-tracking data it was addressed whether L1 and L2 speakers were more likely to fixate the picture of the target than the competitor object, before encountering the noun, in this way demonstrating anticipatory eye-movements driven by the verb. Thus, an anticipatory effect should be visible in terms of more fixations on the animate object for an animate-biasing verb and more fixations on the inanimate object for an inanimate-biasing verb relative to a competitor. To take into account the trajectory of the effect, a growth curve analysis (Mirman et al., 2008) was conducted in R (Dink & Ferguson, 2015; R Core Team, 2017). It was expected that from a 'neutral' baseline, in the current design the picture of the subject/Agent, the probability of looks to the target would increase linearly, reflecting one change in focus, here assessed through the linear orthogonal time polynomial (ot1). For the linear-mixed effects model, the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) was used and to obtain p-values, the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). The model included the factors animacy condition, group and linear time as fixed effects as well as their respective interactions. The random effects structure comprised subjects and items as random intercepts, the interaction between condition and group as by-item slope and the interaction between condition and linear time as by-subject slope. The maximal model (Barr, Levy, Scheepers, & Tily, 2013), including also the interaction with linear time as by-item slope, did not converge. After the maximal model failed to converge, the random effects structure was simplified, and the least maximal models compared by means of the Akaike information criterion (AIC). The model with the lowest AIC value was then selected (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). The factors condition (animate vs. inanimate) and group (L1 vs. L2) were sum contrasted (0.5, -0.5), so the intercept corresponds to the grand mean or average probability that the target was fixated within the critical window across groups and conditions. The dependent variable proportion of looks to target out of looks to target and competitor was transformed into empirical logits (Elog). Table 5.4 shows the output of the model with the formula below.

	Estimate	Std. Error	*t*-value	*p*-value
Intercept (target)	0.4951	0.1347	3.676	0.0005 ***
Condition	0.1511	0.2568	0.588	0.5593
Group	0.2093	0.2282	0.917	0.3627
Linear Time	0.7439	0.2479	3.000	0.0042 **
Condition x Group	0.09	0.4055	0.222	0.8251
Condition x Linear Time	**-1.6741**	**0.4872**	**-3.436**	**0.0012 ****
Group x Linear Time	-0.0697	0.4951	-0.141	0.8886
Condition x Group x Linear Time	0.3386	0.9733	0.348	0.7295
Intercept (condition = animate)	0.5706	0.1867	3.057	0.0034 **
Linear Time	-0.0932	0.2997	-0.311	0.7572
Intercept (condition = inanimate)	0.4195	0.1855	2.262	0.028 *
Linear Time	1.581	0.3896	4.058	0.0002 ***

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Formula: Elog ~ condition*group*ot1 + (1 + condition*group | item) + (1 + condition*ot1 | subject)

Table 5.4: Results of the growth curve analysis for the critical window in Experiment 1. Below the output of the model with sum contrast coding for condition, the results of the follow-up analyses based on the interaction (highlighted) are shown.

The results show a target advantage across groups: Within the critical window, participants were more likely to fixate the target than the competitor. An effect of linear time across conditions and groups indicates that the probability of looks to the target increased linearly. However, the increase was different for the two animacy conditions as signaled by the interaction between condition and linear time. To follow up on this interaction, a linear mixed-effects model with treatment contrast coding for condition was computed and the factor condition re-leveled, to obtain the estimates for the animate and inanimate conditions separately. Again, the aim was the maximal random effects structure, but, after not converging, the model without linear time as by-item slope, as the next best fitting model was selected. Below the output of the model with sum contrasts in Table 5.4, the output of these follow-up analyses is shown. The results show that the probability of looks to target increased linearly for the inanimate-biasing verbs but not for the animate-biasing verbs, resulting in an interaction. The absence of an effect of condition on the intercept term in the group model with sum contrast coding for condition indicates that, for the average time window, the probability that the target was fixated across groups was not significantly different between animacy conditions. There was no effect of group and no group interaction.

Figure 5.3 shows the time course during the critical window, including the empirical and fitted curves. The graph for condition shows that the probability of looks to target was already high at the onset of the critical window for the animate condition and remained more or less stable across the critical window, which possibly reflects an overall preference for animate entities, which is not uncommon for

VW experiments. As seen in the group analyses, the target fixation probability increased in the inanimate condition in a linear fashion. Although there was no significant between-group difference, the graph for group points to a higher probability of looks to the target in the L1 group.

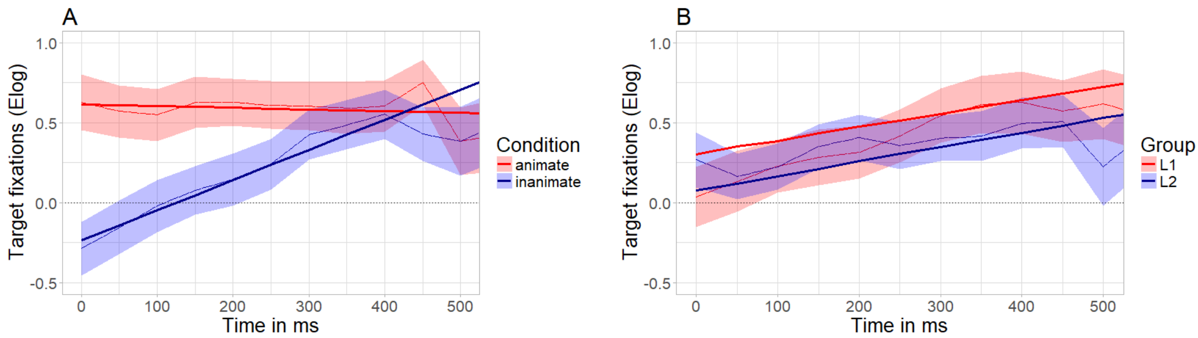


Figure 5.3: Time course showing the probability of looks to target during the critical window in Experiment 1 for (A) condition and (B) group. The curves show the raw data (light) and model prediction (bold).

No between-group differences does not necessarily mean that the groups were exactly the same. To corroborate the conclusion that the groups pattern alike as previously predicted, separate analyses per group were conducted. Note though that, strictly speaking, the missing interaction did not allow the splitting of groups. Figure 5.4 shows the probability of looks to the target per condition during the critical window for each group. Visual inspection points to less certainty in the L2 group. The corresponding model output with treatment contrast coding for condition is shown in Table 5.5 for the L1 group and Table 5.6 for the L2 group. Here the maximal models converged.

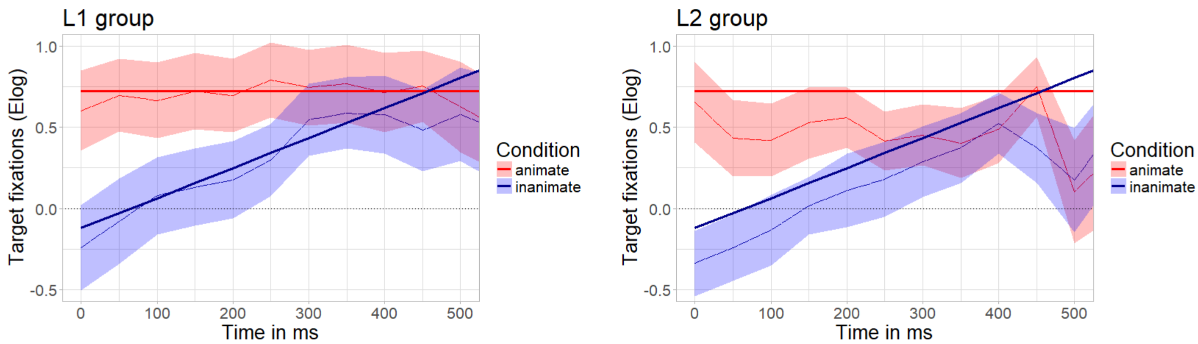


Figure 5.4: Time course showing the probability of looks to target during the critical window in Experiment 1 per group, the L1 group (left) and L2 group (right). The curves show the raw data (light) and model prediction (bold).

The separate analyses show that only the L1 group demonstrated a target advantage in both conditions. The different slopes for the animate- and inanimate-biasing verbs led to an interaction in the L2 but not in the L1 group; however, in both groups the anticipatory effect for the inanimate-biasing verbs developed over time as seen in the group analyses. Hence, whereas an overall target advantage was seen in the L1 group, the effect approached marginal significance for the animate-biasing verbs in the L2 group and only developed over time for the inanimate-biasing verbs. The results from separate analyses thus point to a different timing, presumably also to less certainty, in the L2 group, although the pattern largely resembled that of the L1 group.

L1 group	Estimate	Std. Error	*t*-value	*p*-value
Intercept (condition = animate)	0.7248	0.2443	2.967	0.005 **
Condition	-0.198	0.311	-0.636	0.5281
Linear Time	0.0007	0.6308	0.001	0.9991
Condition x Linear Time	1.5502	1.0239	1.514	0.1389
Intercept (condition = inanimate)	0.5269	0.2447	2.154	0.0369 *
Linear Time	1.5509	0.7470	2.076	0.0457 *
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Formula: Elog ~ condition*ot1 + (1 + condition*ot1 | item) + (1 + condition*ot1 | subject)				

Table 5.5: Results of the growth curve analyses for the L1 group in Experiment 1

L2 group	Estimate	Std. Error	*t*-value	*p*-value
Intercept (condition = animate)	0.4345	0.2546	1.707	0.0956 .
Condition	-0.0737	0.3605	-0.205	0.8393
Linear Time	-0.3019	0.6439	-0.469	0.6425
Condition x Linear Time	2.0598	0.8781	2.346	0.0265 *
Intercept (condition = inanimate)	0.3608	0.2433	1.483	0.1461
Linear Time	1.7579	0.6187	2.841	0.0078 **
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Formula: Elog ~ condition*ot1 + (1 + condition*ot1 | item) + (1 + condition*ot1 | subject)				

Table 5.6: Results of the growth curve analyses for the L2 group in Experiment 1

The growth curve analyses above provided information about the trajectory of the effect (more looks to target vs. competitor) and whether this was different for the two conditions and participant groups. The group model did not show any differences between groups but a target advantage across groups for the critical window. Separate analyses pointed to a later timing in the L2 group. To further explore this finding, a bootstrapped cluster-based permutation analysis (Maris & Oostenveld, 2007) as implemented in the eyetrackingR package (Dink & Ferguson, 2015) was conducted separately for each group to determine the onset of the anticipatory effect. The dependent measure corresponded to the animate AOI fixation proportion within each 50 ms time bin for the 1000 ms time window after verb offset (shifted 200 ms forwards), when taking into account only the two critical AOIs, excluding the picture of the already heard subject/Agent. In this analysis, it was tested in which time bins the curves significantly diverged and the fixation proportions for animate- and inanimate-biasing verbs were significantly different from each other. For better illustration, the graphs in Figure 5.5 show the proportion of looks for the animate AOI after and also shortly before verb offset, to also show potential baseline differences. A paired *t*-test was run on each time bin in the 1000 ms time window after verb offset quantifying the statistical significance of the effect, thus identifying time bins where conditions significantly differed. Adjacent time bins that passed the threshold statistics (α-level = 0.05, two-tailed) were then grouped together into clusters. In a next step, a null distribution was bootstrapped by shuffling the data in each time cluster (here: 1000 times). The observed data were then compared to this bootstrapped distribution. Table 5.7 shows the clusters where an effect was observed, i.e., there was a significant difference between the two animacy conditions. The results indicate that the anticipatory effect emerged at around 50 ms in the L1 group and at around 300 ms in the L2 group. Thus, the cluster-based permutation analysis revealed a delay of around 250 ms in the L2 group.

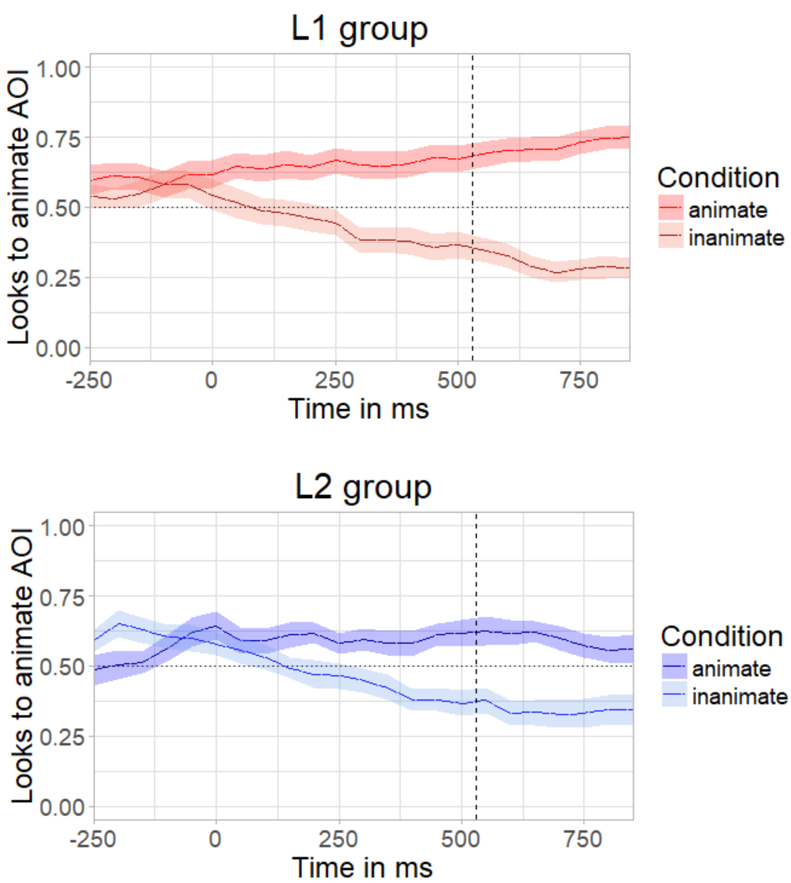


Figure 5.5: Time course showing the proportion of looks for the animate AOI before and after verb offset in Experiment 1. The verb offset shifted 200 ms forwards is aligned to zero, and the dashed vertical line marks the mean onset of the direct object noun.

Finally, note that the results of the permutation analyses should be taken with caution for several reasons: The current design may not have been ideal for testing for the speed of processing. Here, a

Group	Cluster	Direction	SumStatistic	StartTime	EndTime	Probability
L1	1	Positive	91.119	50	1000	0
L2	1	Positive	2.220	200	250	0.212
	2	Positive	40.879	300	1000	0.005

Table 5.7: Results of the cluster-based permutation analysis in Experiment 1. The table shows the start and end times for the clusters identified and the probability of seeing the effect by chance.

baseline condition as in previous studies would actually have been useful, since this way the onset could have been determined separately for the animate and inanimate conditions. Visual inspection of the graphs as well as a separate growth curve analysis point to less certainty for the animate-biasing verbs in the L2 group. In addition, the difference between both conditions after the onset of the direct object noun appears smaller in the L2 group. However, the predicted word in the current design was also the last word of the sentence, thus differences towards the end might result from final sentence interpretation.

5.4 Discussion

Experiment 1 tested whether information on a verb was used to predict a certain feature of an upcoming argument. The source of information was the lexical-semantics of the verb, in particular whether the verb restricted the selection of potentially upcoming nouns to the semantic category animate or inanimate. In the current experimental setting, pictures of an animate and an inanimate object were shown next to the picture of the subject/Agent, which was always displayed below in the center. The visual display was the same for both conditions an item appeared in, but only one of the two upper pictures would show a plausible object completing the sentence depending on the verb's animacy bias.

The results from Experiment 1 showed that the L2 group largely resembled the L1 group when the predictive cue was the lexical-semantics of a verb and the level of prediction the animacy feature of a noun. A growth curve analysis showed no effect of or interaction with group but a target advantage across groups. As expected, the research question could be answered positively: L2 speakers can use the lexical-semantics of verbs to predict the animacy of the upcoming direct object noun. The results also showed that both groups displayed a preference for animate over inanimate objects. A preference for animate objects has also been observed in other studies (e.g., Kamide, Altmann, & Haywood, 2003, p. 139) and presumably results from the fact that animate entities are semantically more salient. For the current data, this was reflected by an interaction between the animacy condition and linear change in time: For the inanimate condition, the likelihood that the target was fixated increased more than for the animate condition. A model testing for a simple effect of linear change in time showed that there was no effect of linear time when the animate condition was taken as the reference level. The eye-tracking results are in line with the finding that the lexical-semantics of verbs is a cue used early by children and, in general, is a very reliable predictive cue for language learners.

To corroborate that the L1 and L2 groups patterned alike, the two groups were further analyzed separately. The separate models indicated that there was a difference in timing, with the L1 group showing a target advantage for the average critical time window in both conditions that increased linearly for the inanimate-biasing verbs, but a marginal effect for the animate condition and only a developing effect in the inanimate condition in the L2 group. To follow up on this difference, the onset of the anticipatory effect was determined for each group. The point in time at which the probability that the animate AOI was fixated significantly diverged between the experimental conditions was different for the groups: Whereas the onset was at around 50 ms after the corrected verb offset in the L1 group, the onset for the L2 group was at around 300 ms. Hence, although the pattern was largely the same and the L2 speakers showed a prediction effect in terms of more/increasing looks to the target compared to a competitor, they appeared to be slightly slower. Another observation was that the L2 speakers were, overall, less certain than the L1 speakers, particularly for the animate-biasing verbs. The results fit in the pattern of results observed by Ito, Pickering, and Corley (2018) described under subsection 4.2.2. Since the authors were interested in the pre-activation of phonological information and, hence, it was important that participants predicted a specific lexical item, cloze probabilities were obtained from separate L1 and L2 groups.[24] The L2 cloze probability was significantly lower than the L1 cloze probability. This may already indicate that L2 expectations were less strong than L1 expectations. The growth curve analyses for the VW experiment, conducted separately for participants and items, showed interactions of experimental condition by language group that indicated a later onset of the anticipatory effect in the L2 relative to the L1 group. An interaction with group on the quadratic term in the by-item analysis indicated that the largest difference between the unrelated baseline condition and the target condition was in the middle of the critical window for the L1 group, but towards the end for the L2 group. Together with the absent English phonological competitor effect in the L2 group, Ito, Pickering, and Corley interpreted their findings in terms of less detailed prediction in L2 processing, probably due to resource limitations. Interestingly, when comparing the time course graphs in the experiment by Ito, Pickering, and Corley (2018, p. 6, Fig. 2), the proportion of looks for the critical object in the target condition rose above 0.5 in the L1 group but remained below it in the L2 group. Like in the current experiment, the L2 speakers in this study seemed less certain throughout. Here, also towards the end of the sentence when the target noun was encountered, the difference in animate AOI fixation proportions appeared smaller in the L2 than in the L1 group, indicating that not only prediction was affected (see also Mitsugi, 2018; van Bergen & Flecken, 2017). In the current design, the predicted word was also the last word of the sentence.

[24]In Experiment 1 the cloze test mainly served another purpose, namely to control for lexical-semantic restrictions, so a verb categorized as animate-biasing would not occur with an inanimate object in a plausible context.

Future experiments may further explore the origin of these subtle differences. The uncertainty could be the result of differences in the lexical representation as argued previously (Hopp, 2013, 2016), here between an L1 and an L2 group; see Dijkgraaf et al. (2017), who found a later onset for bilinguals in both their L1 and L2 compared to monolinguals. Note though that lexical knowledge was tested through a vocabulary list, and trials with verbs unknown to L2 speakers were removed from the analyses of Experiment 1. Nevertheless, it cannot be ruled out that the L1 speakers still differed from the L1 speakers in their lexical retrieval, which was not further tested here.

5.5 Conclusion

Experiment 1 tested the use of lexical-semantic information for prediction, while the level of prediction was systematically controlled for. The results showed that highly proficient Russian L1-German L2 speakers used the lexical-semantics of verbs to predict the animacy of the upcoming direct object noun in German. Within a critical window for an anticipatory effect, participants across groups were more likely to fixate the target than a competitor picture, for example the inanimate noun *Bluse* (blouse) while also presented with the picture of a dark cat after *Die Frau bügelt* (The woman irons) within the article-adjective region *die schwarze* (the black). The results thus confirmed that the L2 group made use of non-grammatical information for prediction, further establishing that the same L2 group tested in Experiment 2 and Experiment 3 can, in general, predict. Nevertheless, follow-up analyses indicated that the anticipatory effect emerged immediately after verb offset in the L1 group but more slowly (although clearly before target noun onset) in the L2 group, who displayed more uncertainty. Hence, if "reduced" is interpreted as slower and/or less certain, the assumption of the RAGE hypothesis does hold, while at the same time the results demonstrate that prediction in L2 processing is possible.

Chapter 6

Experiment 2: Prediction based on verb number marking

6.1 Introduction

Experiment 1 examined the prediction of an upcoming noun's animacy based on the lexical-semantics of the verb in German SVO sentences. Experiment 2 further examined the prediction of the number feature of an upcoming noun, again based on information on the verb. In German, similar to English and Russian, the verb agrees in number with the subject. Here, the verb or, in the following experiment, the copula was placed in front of the subject. In German, this can easily be done by fronting an adverbial phrase, as in the example from Experiment 2 shown in (65). A second adverbial phrase or adjunct (*at noon*) was inserted between the verb and the subject, functioning as the critical window for an anticipatory effect.

(65) Im Briefkasten **ist/sind** am Mittag **eine/mehrere** Zeitung-∅/Zeitungen-**en**.
in the mailbox **is/are** at noon **a/several** newspaper-∅/newspaper-**s**
'There is/are a/several newspaper/newspapers in the mailbox at noon.'

Under the assumption that L2 speakers may have difficulties with inflectional morphology or rely less on grammatical cues than L1 speakers, number as an agreement feature contrasts nicely with the semantic feature examined before. Note though that here the irregular verb *sein* (to be) was used, so there was no affix attached to a verb stem. Nevertheless, the number feature encoded on the irregularly inflected verb still needed to be mapped to the number feature of the subject not yet encountered. The sentence structures used were quite simple and should pose no challenge to the highly proficient L2 speakers, who were further familiar with a flexible word order from their L1. However, as discussed below, verb number marking is probably a less reliable cue than number marking within NPs. Hence, this experiment not only explored prediction based on verb number marking in German L2, but also in German L1 speakers.

6.1.1 Subject-verb agreement

Before turning to the processing of subject-verb agreement, particularly the prediction of a subject's number feature based on verb number marking, it is worthwile to recall how it is realized. Here, the description and the examples given by O'Grady (2008) are used. Consider the examples in (66): The two verbs have different argument dependencies: *exist* requires only one argument, while *ignore* requires two arguments that are placed in front of and after the verb. The verbs are further inflected. The form *ignores* carries the third person singular feature that needs to be matched. While this mapping appears to be easy in (66-a), there are two options in (66-b), hence, one needs to know which argument is the subject; in English this is typically defined by word order.[25] According to O'Grady, the computational system as part of a general sentence building system operates in a linear manner that, in order to minimize working memory costs, tries to resolve dependencies at the first opportunity.

(66) a. Problems exist.
b. Mary ignores Sam.

Things get more complicated in (67), where the subject appears after the verb. In (67-b), the verb only agrees with the first conjunct of the coordinate NP (partial agreement). However, this is only the case for coordinate NPs that follow the subject as demonstrated in (67-c). In German, as shown in (68), both variants are more or less acceptable.

(67) a. There is milk on the table.
b. There is [milk and water] on the table.
c. [Milk and water] are/*is on the table.

(68) a. Es ist Milch auf dem Tisch.
b. Es ist/?sind [Milch und Wasser] auf dem Tisch.
c. [Milch und Wasser] sind/?ist auf dem Tisch.

The examples above show that the number feature of a verb depends on the linear order and that agreeing verbs can precede or follow the subject. Another issue with number marking is that there is a difference between grammatical and notional or real-world number, as discussed, for example, by Riordan et al. (2015). Although often the case, these do not always match as shown in (69-b) in comparison to (69-a).

(69) a. There is a duck on the pond.
b. There is a group of ducks on the pond.

[25] As discussed in the introduction to Experiment 3 (section 7.1), German and Russian, in contrast, are languages with a relatively free word order.

Related to this, Lukyanenko and Fisher (2016) note that one might argue that testing number marking as a predictive cue, see subsection 6.1.2, not only includes grammatical but also notional number, i.e., simply predicting more than one after a plural marking is encountered. However, they also point out that this is highly unlikely also considering that English *are*, for example, precedes or follows second-person subjects in singular and plural and there also exist non-count nouns (e.g., glasses, scissors) that do not reflect real-world number.

In a framework like the Competition Model (MacWhinney, 2001), it is assumed that the order in which cues are acquired is determined by their reliability, with the most reliable cue being acquired first. In English, subject-verb agreement is highly available as a cue, however it is not always also contrastive, see example (70), which makes a cue useful. Only in (70-b) is the agreement cue available and contrastive, since only one noun can agree with the verb. According to MacWhinney (2001), a cue is only reliable if it leads to the right functional choice whenever it is present.

(70) a. The cat chase-**s** the dog.
b. The cat chase-**s** the dogs.

Therefore, subject-verb number agreement and, moreover, its use for prediction is interesting to examine, because it is a grammatical cue that is not always useful. Nevertheless, as further shown below, researchers have considered the possibility that, when a number feature is encountered, it generates an expectation, and one of two studies on English, a morphologically less rich language than German, has provided evidence for prediction based on verb number marking not only in adults but also in children.

Experiments on L2 speakers' comprehension of number agreement show mixed results. Whether they are sensitive to number morphology may depend on whether it is encoded in their L1 (e.g., Chen, Shu, Liu, Zhao, & Ping, 2007; Jiang, 2004), whether it is similarly realized in the L1 and L2 (e.g., Tokowicz & MacWhinney, 2005) and/or on the L2 speakers' proficiency (e.g., Hopp, 2010). For example, whereas the English L1 speakers in Jiang (2004) showed a slow down on the two words after *cabinet* in (71-b) compared to (71-a) in self-paced reading, the advanced Chinese L1-English L2 group showed no significant difference between the two experimental conditions. Chinese lacks grammatical morphology to mark number.

(71) a. The key to the cabinet was rusty from many years of disuse.
b. The key to the cabinets was rusty from many years of disuse.

Experimental findings by Hopp (2010) show that the mastery of subject-verb agreement can be attained, but may be subject to processing load. Crucially, the near-native Russian L1-German L2 speakers he tested patterned together with the L1 group in a self-paced reading experiment and were able to detect subject-verb agreement violations in a speeded acceptability task. Further indication of the mastery of subject-verb agreement in German by Russian natives is provided in a study by Lago and Felser (2018) on agreement attraction that finds similar effects for a German L1 and a highly proficient Russian L1-German L2 group.

Agreement attraction occurs when an attractor, in example (72) below *cabinets*, intervenes between the subject and verb. While in production people are likely to produce a sentence like (72), they also have fewer problems in comprehending such a sentence, although it is ungrammatical (e.g., Pearlmutter, Garnsey, & Bock, 1999).

(72) *The key to the cabinets were rusty from many years of disuse.

In an extensive study, Wagers, Lau, and Phillips (2009) found that agreement attraction only occurs in ungrammatical constructions.[26] They, moreover, showed that agreement attraction was not restricted to prepositional modifiers, but also occurred with relative clause structures like in (73), however only when the head of the relative clause was plural and the subject singular. Hence, agreement attraction also occurs when there is no intervening noun and the attractor is structurally and linearly more distant from the verb than in PP constructions as in (72).

(73) *The musicians who the reviewer praise so highly will probably win a Grammy.

Based on their findings, the authors argue that agreement attraction in comprehension results from difficulties in re-accessing information about the subject at the verb. They provide two possible explanations: One is a cue-based retrieval process, where the number-matching non-subject is the best match when neither of the NPs fully matches. Another is a predictive process, where, after encountering a subject NP, a verb marked with the correct agreement features is expected to follow. If bottom-up features and top-down predicted features then mismatch, cue-based retrieval may become necessary for re-analysis. The following experiment should show whether the parser indeed builds up an expectation when encountering a number feature, here when the verb is presented before the subject and there is only an adverbial in between.

6.1.2 The predictive use of verb number marking

In subsection 4.1.3, the results from an experiment conducted by Lukyanenko and Fisher (2016) on the predictive use of subject-verb number agreement with toddlers and adults were reported that showed an effect of prediction in English-speaking children at the age of three years. An overview of their experimental design with one example in all conditions is given in Table 6.1. Considering the predictive use of number marking in young children, it is therefore more surprising that Riordan et al. (2015) found

[26] According to them, additional processing costs that were detected in structures like (71-b) rather result from costs associated with processing plural nouns.

target	condition	trial	[...] verb	CRITICAL WINDOW	subject
Singular	experimental:	informative	Where is	the good	apple?
	one apple, two cookies	uninformative	Can you find	the good	apple?
Singular	control:	informative	Where is	the good	apple?
	one apple, one cookie	uninformative	Can you find	the good	apple?
Plural	experimental:	informative	Where are	the good	cookies?
	one apple, two cookies	uninformative	Can you find	the good	cookies?
Plural	control:	informative	Where are	the good	cookies?
	two apples, two cookies	uninformative	Can you find	the good	cookies?

Table 6.1: Experimental design in Lukyanenko and Fisher (2016)

no effect, that is, no difference between same or uninformative and different number trials, for adult English L1 speakers.

Riordan et al. presented participants with two pictures of different objects and sentences of the form *There/Where is/are [determiner] [noun].* In the first two experiments, the sentences appeared in two conditions, a definite determiner condition with a single number cue on the copula (e.g., *There is/are the lion/s.*) and an indefinite determiner condition with a number cue on the copula and the determiner (e.g., *There is/are a/some lion/s.*), whereby in the second experiment the task was to click on the picture that was mentioned in the sentence and to do so as quickly as possible to speed up processing. In a third experiment, the authors added more conditions including different sentence structures, with ten sentence sets in total, as shown in the overview in Table 6.2.

	condition					
Experiments 1–2	Definite, Singular	There/Where	is	the	lion	
	Definite, Plural	There/Where	are	the	lions	
	Indefinite, Singular	There/Where	is	a	lion	
	Indefinite, Plural	There/Where	are	some	lions	
Experiment 3	Singular	Is	there	a	dog	
	Plural	Are	there	some	dogs	
	Singular		Does	the	dog	have brown fur
	Plural		Do	the	dogs	have brown fur
	Singular			That	dog	is black
	Plural			Those	dogs	are black

Table 6.2: Experimental design in Riordan et al. (2015)

The authors measured the reaction time, here defined as the latency of the first saccade or fixation marking the beginning of an uninterrupted series of fixations on the target picture, from the onset of the copula onwards. All on-target trials, trials with a saccade or fixation on the target 200 ms prior to copula onset and trials with reaction times that occurred later than 700 ms after noun onset were removed from the analysis. The authors do not see a lack of power as responsible for the null findings. However, also taking into account the short time window for an effect of prediction (the mean length of the determiner segment in Experiments 1 and 2 was 151 ms, ranging from 50 ms to 275 ms), the experimental design might have affected the outcome. Based on their results, Riordan et al. argue that number morphology only has a low cue validity in adults and the more experienced a language user is, the less reliable number morphology may become considering the many other potential continuations (e.g., *There is a group of [noun$_{sg.}$])*.[27] As a result, they may resort to a form of "good-enough" processing (Riordan et al., 2015, p. 8).

Findings by Hopp (2012) show that Russian L1-German L2 speakers were able to predict an upcoming noun based on number inflection on an article or adjective. However, number marking on verbs, as shown above, is probably less reliable than number marking within NPs. Following the argumentation of Riordan et al. (2015), L2 speakers actually may be better predictors because they typically have less experience with the language and, thus, have encountered fewer examples where the verb number cue has been misleading. In addition to its relevance for the question whether prediction is selectively limited, including L2 speakers was also a way of testing this ad-hoc hypothesis.

The experimental design of Experiment 2 differed from previous studies that compared informative and uninformative trials. The current design was similar to the one in Experiment 1, where participants were presented with a visual display showing three AOIs, while listening to a spoken sentence, here either in a singular (*is*) or plural (*are*) condition. Instead of analyzing the time participants need to shift their gaze towards the target, a growth curve analysis as for Experiment 1 was conducted. Experiment 2 addressed the following research question:

> RQ: Do L1 and L2 speakers of German use verb number marking to predict the number feature of the upcoming subject?

If the groups tested used the number cue predictively, they should be more likely to fixate a single object picture after *ist* (is) than after *sind* (are) and to fixate a multiple object picture after *sind* (are) than after *ist* (is) before the onset of a quantifier. Note that this experiment was explorative and the predictions less clear for both groups. However, based on the findings from Lukyanenko and Fisher (2016) for English, who found a predictive use of verb number marking in adults and children, a predictive use of verb number marking in German, a morphologically richer language than English, appeared more likely than no effect as reported by Riordan et al. (2015). Hopp (2012) could show that adult German L1 and English

[27] Note that this stands in contrast to what has been proposed by MacWhinney (2001) regarding the acquisition of cues based on reliability. Following Riordan et al., an acquired cue would be somehow 'unlearned.'

L1-German L2 speakers were able to use number agreement within NPs in German. In his study, the participants used number marking on an article/adjective to anticipate the upcoming noun.

6.2 Methods

6.2.1 Participants, apparatus & experimental procedure

Experiment 2 was conducted together with Experiment 1 and Experiment 3 in one experimental session. The experimental items appeared within the same presentation list. For information about the participants, the apparatus and experimental procedure, see subsection 5.2.1, subsection 5.2.3, and subsection 5.2.4, respectively.

6.2.2 Design & materials

6.2.2.1 Eye-tracking materials

Twenty-eight items in two experimental conditions, resulting in 56 sentences overall, were constructed. An example is given in Table 6.3. All experimental sentences included the third person form of the verb *sein* (to be), marking either (a) singular or (b) plural, because the use of other verbs leads to ambiguity in the plural condition. In German, the third person plural matches the infinitive and can initially also be interpreted as a gerund form.[28] The critical window contained an adverbial with a length of two to three syllables. This region was cross-spliced, so it was identical in both conditions. The sentences were recorded in a sound attenuated room and spoken at a normal speaking rate by the same female native speaker of German who read the sentences for Experiment 1.

	location	verb	CRITICAL WINDOW	quantifier + noun
(a) singular	Im Briefkasten In the mailbox	**ist** is	am Mittag at noon	eine Zeitung. a newspaper.
(b) plural	Im Briefkasten In the mailbox	**sind** are		mehrere Zeitungen. several newspapers.
'There is/are a/several newspaper/newspapers in the mailbox at noon.'				

Table 6.3: Auditory stimulus in Experiment 2. Each item appeared in two conditions, with a singular verb (a) or plural verb (b). The critical window was cross-spliced.

Participants were presented with three pictures, a picture of the location mentioned clause-initially at the bottom center of the screen and the singular and plural versions of the object either to the left

[28]Example: *Am Baum Hängen macht mir Spaß.* – Hanging from the tree is fun to me.

or right side above; see Figure 6.1 for an example. The position of the respective target picture was counterbalanced. As for Experiment 1, the pictures were taken from the MultiPic database (Dunabeitia et al., 2018) complemented by new drawings and/or drawings taken from https://openclipart.org; in the case of the singular and plural objects, the pictures were edited accordingly so they showed either one single object or three identical smaller versions (only exception: the singular version for child/children showed a girl and the plural version two girls and a boy). The items were equally distributed across two lists using a Latin square design.

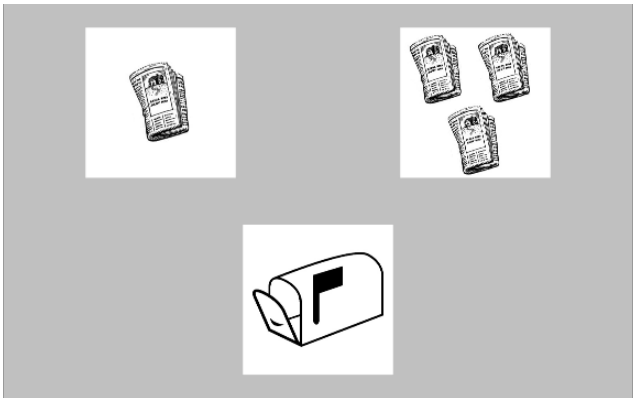


Figure 6.1: Visual display in Experiment 2. The picture in the bottom center shows the location and the two upper pictures the singular and plural versions of the same object.

Ten additional fillers also included sentences containing regular verbs as in (74-a) or sentences in which grammatical and notional number mismatched as in (74-b), to avoid participants developing a strategy based on *ist/sind* (is/are). Moreover, two other experimental sets (see Experiment 1 and Experiment 3), consisting of 24 items and 28 items, and further 18 fillers, including six with singular and plural manipulation but showing four pictures, were included in the experiment. All experimental sentences and fillers are listed in Appendix B under section B.2 and section B.4.

(74) a. Hinter dem Haus bellt am Abend ein Hund.
behind the house barks in the evening a dog
'A dog barks behind the house in the evening.'

b. Auf dem Baum sitzt mittags eine Schar Vögel.
on the tree sits at noon a flight birds
'A flight of birds sits on the tree at noon.'

In (75), two possible Russian translations for the example item above are given. In Russian, verbs are also marked for number, but for the present tense the verb is not overtly realized.

(75) a. W potschtowom jaschtschikje __ w poldjen' odna/njeskol'ko
in the mailbox is/are at noon **one/several**
gasjet-**a**/gasjet-∅.
newspaper-∅/newspaper-**s**

b. W potschtowom jaschtschikje jest' w poldjen' odna gasjeta.
in the mailbox is at noon one newspaper

Another option is to insert the verb *jest'* without number marking, which, in the current example, indicates that the mailbox is not empty at noon. Hence, verb number marking exists in Russian, so the L2 speakers tested were familiar with it, but is not realized in the construction under investigation.

6.2.2.2 Offline test materials

In the additional offline test, the L2 speakers were presented with the same, but slightly shortened materials, see (76), and had to select the correct completion. All participants received the items from the other list they had not already been exposed to in the eye-tracking experiment.

(76) Im Briefkasten ist ('In the mailbox is')

a. mehrere Zeitungen ('several newspapers').

b. eine Zeitung ('a newspaper').

Participants completed the offline questionnaire after the eye-tracking experiment. The offline questionnaire consisted of a forced-choice sentence completion task and, for Experiment 1, a vocabulary list including the verbs from the respective list the participant was assigned to. The 28 sentences testing participants' knowledge about subject-verb agreement were mixed with 28 sentences testing their knowledge about case marking (see Experiment 3) and ten fillers from the eye-tracking experiment. For the questionnaire, Google forms were used. Depending on the WiFi connection in the test room, the participants completed the offline test on the web or received the paper version. They were further encouraged to write down vocabulary they were not familiar with.

6.2.3 Procedure: Data pre-processing

For the analyses of the eye-tracking data, four items that included the adverbial *immer* (always) were excluded, because the critical time window for an anticipatory effect was much shorter than for all other items. This left 24 items for the analyses. The third block for one of the L1 speakers and the first block for one of the L2 speakers were excluded, for the same reasons as in Experiment 1, as well as one trial that was accidentally skipped by one participant in the L1 group, removing 2% of the data in the L1 and 1% in the L2 group. As in Experiment 1, participants could fixate either one of three AOIs or the background of the screen. Background looks were treated as trackloss (within the critical window 4% for the L1 and 6% for the L2 group). It was expected that participants would fixate the AOI displaying the location mentioned in the beginning of the sentence upon its mentioning, before moving their eyes towards the critical AOIs, the singular AOI or plural AOI, when encountering the verb. Figure 6.2 shows an overview of the time course with the offset of the verb aligned to zero per group and condition. As can be seen in the graphs, the proportion of looks towards the location (grey line) started to increase in all groups and in both conditions. The graphs further show a clear baseline effect: Participants in both groups displayed an overall preference for the plural AOI right from the start.

For the statistical analyses, only fixations on the singular and plural AOIs within the critical window, starting 200 ms after the offset of the verb and ending 200 ms after the offset of the adverbial and the onset of the subject, were considered. The critical window for the remaining data set had an average length of 528 ms (SD: 57.21). Hence, the critical time window in Experiment 2 had a similar length to the critical window in Experiment 1. For the analyses, the exact on- and offset information determined in Praat (Boersma, 2001) was used. Data points were aggregated into 50 ms time bins. In the L1 group, 77% of all On-AOI fixations within the critical window included trials with fixations on the singular or plural AOI, and in the L2 group 79%. The amount of discarded data was similarly distributed across the singular and plural conditions. For all analyses and previous data preparation the eyetrackingR package (Dink & Ferguson, 2015) was used.

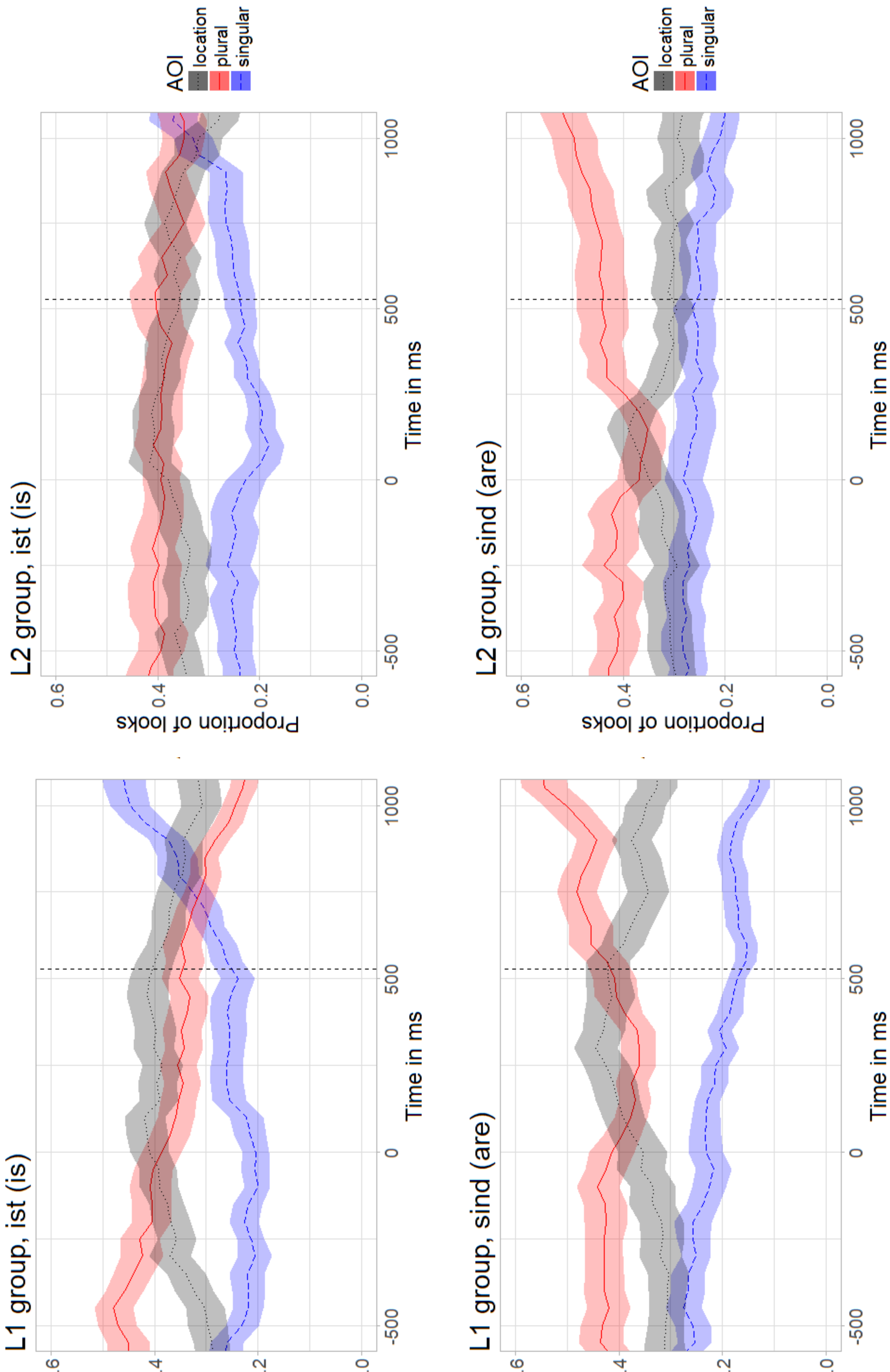


Figure 6.2: Proportion of looks for all AOIs in Experiment 2 for the L1 group (left) and L2 group (right). The plots show the proportions for the singular verb (top) and the plural verb (bottom); the verb offset shifted 200 ms forwards is aligned to zero, and the dashed vertical line marks the mean onset of the subject.

6.3 Results

For the behavioral data, i.e., the response accuracy for the statements following half of the sentences in the eye-tracking experiment, see subsection 5.3.1.

6.3.1 Offline data

The participants included in the analysis had no problems in choosing the correct singular or plural form in the additional offline test but scored at ceiling. Their mean accuracy was 0.998% with only one incorrect response. Hence, the L2 group displayed no variance for the use of number marking. The outcome of the offline test further confirms the assumption that the investigated structures were unproblematic for the highly proficient L2 speakers.

6.3.2 Eye-tracking data

As mentioned in the description of the data pre-processing, participants showed a preference for the plural AOI throughout the trials. One explanation for the preference for the plural AOI, in addition to its visual salience, could be that participants in the singular condition fixated one of the three objects in the multiple object picture instead of the singular AOI. Hence, the growth curve analysis tested whether participants over time were less likely to fixate the singular AOI after *sind* (are) than after *ist* (is). Participants should not have fixated the single object picture after *sind* (are) as this would have been clearly implausible.

The dependent variable, the proportions of looks to the singular AOI out of all singular AOI and plural AOI fixations within the critical window, was transformed into empirical logits (Elog). The mixed-effects model included the interaction between condition (singular vs. plural), group (L1 vs. L2) and linear time as fixed effects, as well as their respective interactions. After the maximal model failed to converge, the random effects structure was simplified. After comparing the least maximal models that converged, the final model selected included subjects and items as random intercepts and the interaction between condition and group as by-item slope, as well as the interaction between condition and linear time as by-subject slope. Note that an effect of verb number marking in this model should show up as an effect of condition, signaling fewer looks to the singular AOI for the plural condition, or, if developing over time, as an effect of linear time or interaction with linear time. The output of the model with treatment contrast coding for condition and sum contrast coding for group (0.5, -0.5) is shown in Table 6.4 with the formula below. The intercept in this model corresponds to the average probability that the singular AOI was fixated in the singular condition for both the L1 and the L2 group.

Over time, participants were more likely to fixate the singular AOI after *ist* (is) than after *sind* (are), as indicated by the effect of linear time and the respective interaction between condition and linear time in the model output. To further explore the interaction between condition and linear time, the factor condition was re-leveled, so the plural condition became the reference level. This reveals that the slope was different for the two conditions with no increase (but also no significant decrease) in the probability of looks to the singular AOI in the plural condition. The overall preference for the plural AOI, which was clearly visible in the graphs, can be seen in the output of the growth curve analyses as an effect at the intercepts: In both conditions, participants were less likely to fixate the singular AOI than the plural AOI. The effect of linear time, however, may indicate that across groups an effect of number emerged. However, one word of caution is needed: The effect of linear time showed up when groups were collapsed, which also increased statistical power. A model with treatment contrast coding for group showed that this effect was mostly carried by the L2 group (for the model output, see Table B.5 in Appendix B). For a visualization, see Figure 6.3. Looking at the graphs, it also becomes evident that the groups differed at the onset of the critical window, with the L2 group showing a higher probability of looks to the singular AOI in the plural compared to the singular condition at the onset. Whereas in the L1 group the curves at least visibly start diverging, although this development stagnates later on, the curves in the L2 group cross each other towards the end of the critical window.

	Estimate	Std. Error	*t*-value	*p*-value
Intercept (condition = singular)	-0.4786	0.1734	-2.760	0.0075 **
Condition (plural)	-0.1045	0.2666	-0.392	0.6963
Group	0.1762	0.3293	0.535	0.5944
Linear time	0.6608	0.2991	2.209	0.0319 *
Condition x Group	-0.3613	0.5249	-0.688	0.4937
Condition x Linear time	**-0.9599**	**0.4343**	**-2.210**	**0.0317 ***
Group x Linear time	-0.4253	0.5982	-0.711	0.4805
Condition x Group x Linear time	0.5850	0.8685	0.674	0.5037
Intercept (condition = plural)	-0.5831	0.1653	-3.526	0.0008 ***
Group	-0.1851	0.3232	-0.573	0.569
Linear time	-0.2991	0.2646	-1.130	0.2637
Group x Linear time	0.1597	0.5292	0.302	0.764

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Formula: Elog ~ condition*group*ot1 + (1 + condition*group | item) + (1 + condition*ot1 | subject)

Table 6.4: Results of the growth curve analysis for the critical window in Experiment 2. Below the output of the model with the singular condition as reference level, the results of the follow-up analysis based on the interaction (highlighted) are shown.

To follow up on the observation that an effect of condition only emerged over time, the same bootstrapped cluster-based permutation analysis as for Experiment 1 was conducted, which could

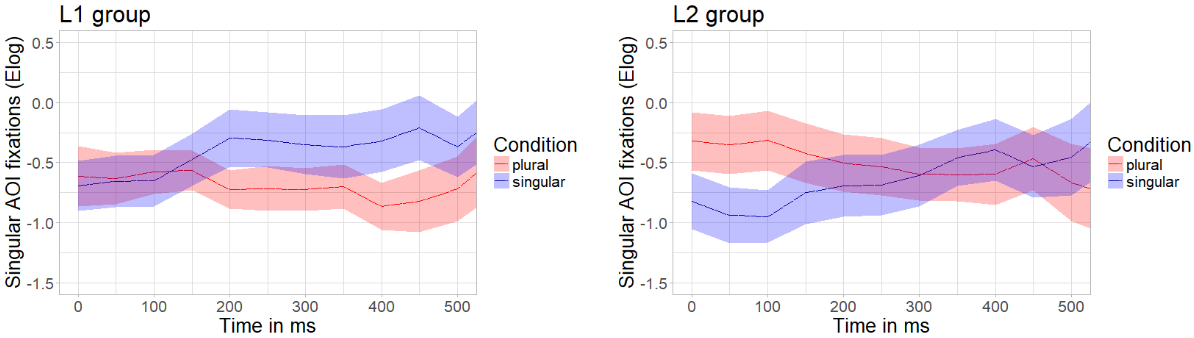


Figure 6.3: Time course showing the probability of looks to the singular AOI during the critical window in Experiment 2 per group, the L1 group (left) and L2 group (right)

determine when the two curves actually diverged, see Figure 6.4; for a description of the procedure, see subsection 5.3.2. The onset of divergence for the L1 group was shortly after the mean length of the critical window, hence only with the onset of the quantifier did a clear difference between *ist* (is) and *sind* (are) emerge. The same time window as for Experiment 1 was used, that is, 1000 ms after the onset of the critical window. For the L2 group, the analysis detected no cluster within this time window. Only after adding 250 ms[29] did the analysis detect a cluster, however the proportion of looks to the singular AOI in the singular condition was not significantly different from the plural condition, see Table 6.5.

Group	Cluster	Direction	SumStatistic	StartTime	EndTime	Probability
L1	1	Positive	27.1442	600	1000	0.006
L2	1	Positive	9.6303	1000	1250	0.084

Table 6.5: Results of the cluster-based permutation analysis in Experiment 2. The table shows the start and end times for the clusters identified and the probability of seeing the effect by chance.

This additional analysis showed that the effect of verb number marking only developed over time and there was no clear effect until the onset of the quantifier. Moreover, the L2 speakers showed no significant difference between conditions within the first 1250 ms. Instead, an effect of verb number marking further developed over time. It should be noted that the probability of looks to the singular AOI in the singular condition did not rise above chance level within the critical window for an anticipatory effect nor shortly after. Although the results have to be taken with caution, it appears safe to conclude that number marking on the verb facilitated later information integration in the L1 group but probably not in the L2 group.

[29]This corresponds to the delay observed in Experiment 1.

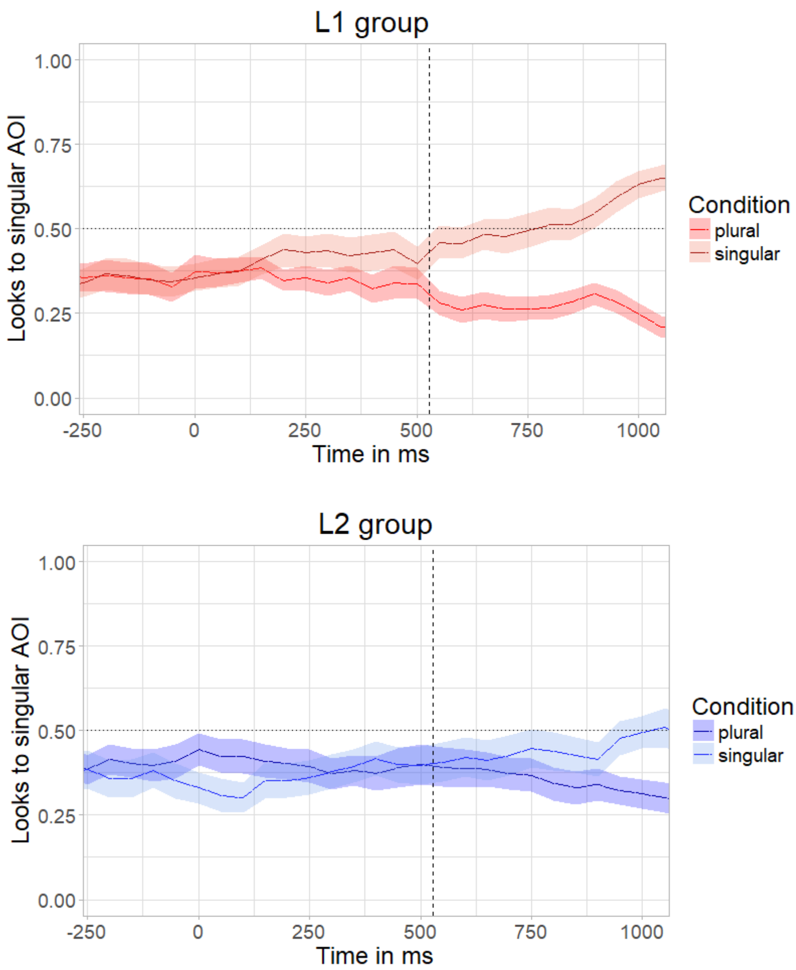


Figure 6.4: Time course showing the proportion of looks for the singular AOI before and after verb offset in Experiment 2; the adverb onset shifted 200 ms forwards is aligned to zero, and the dashed vertical line marks the mean onset of the quantifier+noun window.

6.4 Discussion

Experiments 1 and Experiment 2 tested how information on the verb was used to predict a certain feature of an upcoming noun, in Experiment 1 the semantic feature animacy based on the lexical-semantics of the verb, and here the number feature of the upcoming subject based on number encoded on the verb. The participants were presented with a picture of a location at the bottom center of the screen and a single object and multiple object picture to the left and right above. The descriptive data showed an overall preference for the plural AOI, probably due to the visual salience of the multiple object picture, already starting at sentence onset. A growth curve analysis then tested whether participants were less likely to fixate the single object picture/the singular AOI after *sind* (are) than after *ist* (is).

The results of the growth curve analysis showed that across groups the probability that the singular AOI was fixated increased over the course of the critical window for the singular condition *ist* (is), as shown by an effect of linear time. However, there was no effect of condition but only an interaction between condition and linear time, indicating a different development in the plural condition *sind* (are). Further analyses showed that the effects were only observed across groups. An additional analysis as also conducted for Experiment 1 to determine the onset of the effect confirmed that the two conditions did not significantly differ until into the post-critical window. As soon as the quantifier was encountered, the L1 but not the L2 group showed a clear difference between the singular and plural conditions. In the L2 group, the effect of number marking further developed over time. One explanation for this is given below. Hence, unlike in Experiment 1, there was no anticipatory effect: Participants showed no clear preference for the target over the competitor picture before the target was mentioned, here the subject that agrees in number with the verb. Considering the findings from Riordan et al. (2015), who concluded that verb number marking in English only has low cue validity and, therefore, may not be used predictively by experienced language users, the results of Experiment 2 also add to existing research by testing verb number marking in German not only in L1 but also in (less experienced) L2 speakers. I propose that further research is needed to draw safe conclusions regarding the reliability/validity of a number cue. There is reason to assume that the lack of an anticipatory effect in the current design was, at least partly, caused by the experimental design.

The results of Experiment 2 were confounded by the salience of the plural AOI, which led to a baseline effect and an overall high probability of plural AOI fixations. Therefore, any conclusion can only be based on the observation that participants across groups, over time but before subject onset, were more likely to fixate the singular AOI in the singular than in the plural condition. An effect for the plural condition was masked by the baseline effect. The intercept values were not informative, since in both conditions participants were more likely to fixate the plural AOI. There are several explanations why the plural picture appeared more attractive. First, it included more information. Being aware of the problem with plural pictures, Lukyanenko and Fisher (2016) added a control condition to their experiment, because previous work had pointed to a preference for plural pictures in children. Their results still showed that children's higher probability for distractor-to-target shifts in informative trials tended to be more pronounced when the target was plural. It was less possible to control for the salience of the plural picture in the current design without introducing a further confound, for example by presenting another plural picture that would not fit in the context. Second, one may argue that the plural picture always displayed the singular as well: Three newspapers are at the same time three individual newspapers. Thus, participants could have fixated one of these three instead of the single object picture, especially since the critical AOIs displayed the same objects (see also Robertson et al., 2012). Hence, it cannot be ruled out that some participants (strategically) fixated the multiple object picture and only towards the end of the sentence used the bottom-up information to fixate the target picture. Especially in the L2 group a strategic fixation of the multiple object picture cannot be ruled out. Even in the beginning of the post-critical window there was no significant difference between conditions in the L2 group.

Another finding, visible in the proportion graphs for all AOIs, was that the participants still were distracted by the AOI displaying the location mentioned in the beginning of the sentence. The experimental sentences might not have been optimal insofar as an adverbial was topicalized, resulting in a marked sentence structure. Moreover, a further reason why the location was still attended to could be that it remained relevant. Note that for the example item the mailbox contains the newspaper/newspapers. The other items were similarly constructed. Apart from the problems arising due to the experimental design, eye-tracking might be less suited to test for the comprehension of number marking, a point I will return to in the general discussion (section 9.2).

6.5 Conclusion

Experiment 2 tested the use of a verb number cue to predict the upcoming subject. Although there was some indication that L1 and L2 speakers of German were more likely to fixate the singular AOI (one newspaper) after a sentence like *Im Briefkasten ist* (In the mailbox is) than after the same sentence with *sind* (are) during the critical window *am Mittag* (at noon), the two conditions most clearly differed only after the onset of the subject. There was also a difference between groups, with the L1 group showing a difference between conditions after subject onset, interpreted as a facilitative effect, and no immediate difference for the L2 group. The interpretation of the results was complicated by a baseline effect: Throughout the trials, participants showed a preference for the plural AOI. The experimental design might have enhanced this preference. It could be that participants, maybe strategically, fixated one object in the multiple object picture/plural AOI. It remains a subject for future research whether verb number marking functions as a predictive cue in German.

Chapter 7

Experiment 3: Prediction based on case marking

7.1 Introduction

It is essential for successful language processing to find out who did what to whom, i.e., to have a successful mapping between sentence and event structure, which requires the correct assignment of thematic roles. The thematic roles of the one or more arguments that a predicate takes encode the semantic role (Agent, Patient, etc.). The assignment depends on syntax, whether the argument is the subject, direct or indirect object or oblique, as well as on semantics, like semantic features of the argument and the lexical-semantics of the verb (Becker, 2017). The next paragraphs will show that in a language like German, in active sentences subjects receive nominative case, direct objects accusative case and indirect objects dative case, thus the grammatical function is encoded morphologically, which, in turn, allows for a relatively free word order. However, not every word order variation is equally acceptable and, potentially, equally easy to process (e.g., Dröge, Fleischer, Schlesewsky, & Bornkessel-Schlesewsky, 2016). As will be shown below, some linearizations are more preferred. In the context of the following experiment on the use of case marking to predict an upcoming thematic role, the interest particularly lies in the possible alternation of the dative/indirect object/Recipient and accusative/direct object/Theme in German main clauses. In addition, I will also refer to languages that have been examined in previous studies, in particular Russian, the L1 of the L2 learners tested within this thesis.

7.1.1 Thematic role assignment and word order variation

In the face of missing consensus as to what thematic roles actually are, Dowty (1991) calls them "creatures of the syntax-semantics interface" (p. 548). He suggests only discriminating between Proto-Agents and Proto-Patients, while Primus (1999) also suggests a Proto-Recipient. Another, similar approach is to discriminate between semantic macro-roles as done by van Valin and LaPolla (1997). They discriminate between an Actor and an Undergoer, whereby the Actor is ranked higher in the argument hierarchy. In the context of the following experiment, I will use the traditional terms Agent and Theme as they are described informally by Haegeman (1996) and the term Recipient as described by Primus (2012), which is sometimes used interchangeably with Benefactive or Goal:

- **Agent/Actor:** the one who intentionally initiates the action expressed by the predicate
- **Theme:** the person or thing moved by the action expressed by the predicate
- **Recipient:** receiver of an entity or information for predicates that denote a transfer of possession or information

The semantic relationship between the predicate and its arguments is the special focus of an fMRI study by Bornkessel, Zysset, Friederici, von Cramon, and Schlesewsky (2005), where they tested simple transitive sentences in German. Typically, the Agent/Actor is the subject and is placed in clause-initial position, so its linear position is in alignment with the position in the argument hierarchy. Based on previous ERP studies, the authors assumed that there were two pathways available to the parser, a morphological one and a positional one, whereby the parser would draw upon the first one when case is unambiguously marked. Their main finding was that different neural regions in the brain were particularly sensitive to morphological information and the syntactic realization of the verb-based argument hierarchy on the one hand and to the linearization of hierarchical linguistic information on the other. Interesting for the following experiment, the authors speculate that there might be cross-linguistic differences regarding the specialization of components, mentioning German, Russian and Japanese, in contrast to English, which rely more on morphological case and less on linear order.

In languages with a rich case marking system like German, Russian and Japanese, word order is typically flexible. Case marking and word order can both be used to distinguish between thematic roles and, thus, often represent alternatives to each other (e.g., Blake, 1997). In the following experiment, surface word order or argument linearization was manipulated within the German middle field. The middle field corresponds to the sentence segment starting after the finite verb in German main clauses as in (77) and, as in (78), from the complementizer/conjunction up to the verb in the end position/particle in coordinate or subordinate clauses (see, e.g., Haider, 2017).

(77) Der Vater gibt (‘The father gives’)

a. dem Baby den Schnuller.
[the baby]-*Dat* [the pacifier]-*Acc*
‘the pacifier to the baby.’

b. den Schnuller dem Baby.
[the pacifier]-*Acc* [the baby]-*Dat*

(78) Die Mutter sagt, (‘The mother says’)

a. dass der Vater dem Baby den Schnuller gab.
that [the father]-*Nom* [the baby]-*Dat* [the pacifier]-*Acc* gave
‘that the father gave the pacifier to the baby.’

b. dass der Vater den Schnuller dem Baby gab.
that [the father]-*Nom* [the pacifier]-*Acc* [the baby]-*Dat* gave

c. dass dem Baby der Vater den Schnuller gab.
that [the baby]-*Dat* [the father]-*Nom* [the pacifier]-*Acc* gave

All orders in (77) and (78) are grammatical but they vary in acceptability. For the German middle field, Ellsiepen and Bader (2018), based on previous work (e.g., Jacobs, 1988; Lenerz, 1977; Müller, 1999; Uszkoreit, 1986), discriminate between six different surface constraints, as shown in Table 7.1. Note that the term *acceptability* describes how a speaker perceives a linguistic stimulus, which can be different from grammaticality. In four acceptability judgment experiments, the authors aimed to tease apart the influence of these surface constraints, which are often confounded. The results demonstrated that the syntactic constraint ‘dative > accusative’ was weighted higher than the lexical-semantic constraint ‘Recipient/Goal/Benefactive > Theme,’ however the latter directly followed the former as the lowest ranked constraint behind all others shown in the table below, with ‘nominative > non-nominative (accusative)’ being the highest ranked constraint.

1. Lexical-semantic constraints	2. Syntactic constraints	3. Discourse constraints
a. animate > inanimate b. Agent > non-Agent c. **Recipient/Goal/Benefactive > Theme**	a. nominative > non-nominative b. **dative > accusative**	a. definite > indefinite

Table 7.1: Argument linearization in German: Lexical-semantic, syntactic and discourse constraints. Highlighted in bold are the constraints relevant in Experiment 3.

Previous experiments in German have tested the predictive use of case marking via object topicalization (Henry et al., 2017; Hopp, 2015; Kamide, Scheepers, & Altmann, 2003), where the accusative object was fronted and placed in front of the finite verb, in the so called prefield. The experimental manipulation in Experiment 3 included the switching of dative object/Recipient and accusative object/Theme within the middle field, which was expected to be less marked than object topicalization. Higher ranked constraints like ‘nominative > non-nominative’ and ‘Agent > non-Agent’ were respected in Experiment 3.

The constructions under investigation in Experiment 3 were similar to the ones in (77), where the direct/accusative object and indirect/dative object are swapped, but here the thematic roles were changed. Whereas *the baby* is still the dative object and Recipient in (77-b), it became the Theme in (79-b). If languages mark dative case, for verbs like *give* that take three arguments the dative object will typically encode the Recipient (Primus, 2012, p. 55). In English, dative alternations appear to be more frequent than in German and can be realized either as a double object construction (*the baby the pacifier*) or as a prepositional object (PO) or indirect object construction (*the pacifier to the baby*).

(79) Der Vater gibt ('The father gives')

a. dem Baby den Schnuller.
[the baby]-*Dat* [the pacifier]-*Acc*
'the pacifier to the baby.'

b. das Baby der Mutter.
[the baby]-*Acc* [the mother]-*Dat*
'the baby to the mother.'

Whereas English is considered to be mixed, meaning that ditransitive verbs can occur with double object or prepositional object/indirect object constructions, German and Russian are considered to be languages with indirect object constructions, where the Recipient has a different marking than the Theme (see Haspelmath, 2013).[30]

How do people process sentences with different word orders or sentences that somehow violate constraints in real-time? In Russian, the surface word order is also flexible but highly depends on the discourse context (e.g., Cubberley, 2002; Sekerina, 2003): Already given information is placed in front of new information. Sekerina (1997, 2003) could show that, when sentences were presented in null contexts (no restricting discourse), scrambling resulted in longer reading times, providing evidence for higher processing load compared to unscrambled sentences.[31] A higher processing load associated with scrambling or word order variation in the German middle field appears more difficult to detect (Bader, 1994). However, ERP studies (e.g., Dröge et al., 2016; Rösler, Pechmann, Streb, Röder, & Hennighausen, 1998; Schlesewsky, Bornkessel, & Frisch, 2003) could find a left-anterior negativity elicited by non-canonical structures. Schlesewsky et al. (2003) tested German sentences like (80), where (80-a) is assumed to be non-canonical or marked – and a case of scrambling – but (80-b) is considered canonical because pronouns typically precede non-pronominal arguments, although in (80-b) movement is also involved (see footnote).

[30] Note that in German dative alternations can also be realized as a PO construction, although rather infrequently. Kholodova and Allen (in press) found that these constructions (e.g., *Der Bauer verkaufte die Ente an die Dame* – The farmer sold the duck to the lady) are less restricted than typically assumed, but that their occurrence is conditioned by the verb used and the modality (production vs. comprehension).

[31] In Russian, scrambling is less restricted than in other languages that allow scrambling. The term *scrambling* refers to the re-ordering of arguments in a sentence. For German, it is confined to the middle field. The term has been used in different contexts and it is unclear whether scrambling is the result of a derivational process or movement, or whether there are just sets of individually available serialization patterns that are all base-generated (e.g., Haider, 2017). I thus prefer the use of the more theory-neutral term *word order variation* throughout.

(80) a. Dann hat dem Jäger der Lehrer den Roman gegeben.
then has [the hunter]-*Dat* [the teacher]-*Nom* [the novel]-*Acc* given
'Then the teacher gave the novel to the hunter.'

b. Dann hat ihm der Lehrer den Roman gegeben.
then has him-*Dat* [the teacher]-*Nom* [the novel]-*Acc* given
'Then the teacher gave him the novel.'

The results showed a negativity for non-canonical word orders like (80-a) but not for (80-b), which the authors took as an indication that this component is not the result of working-memory load (cf. Rösler et al., 1998), but rather reflects a local syntactic mismatch. More recently, the authors interpreted this negativity in terms of a prediction mismatch (Dröge et al., 2016). In a VW experiment by Mitsugi and MacWhinney (2016) that tested the use of Japanese case markers for prediction, no difference between the canonical structure 'subject/nominative > indirect object/dative > direct object/accusative > verb' and the scrambled structure was observed. Although scrambled structures in Japanese have a low frequency, other experiments (Yamashita, 1997) have also detected no differences in reading times between the canonical and a non-canonical order.

7.1.2 The predictive use of case marking

In chapter 4, two VW experiments in Japanese (Kamide, Altmann, & Haywood, 2003; Mitsugi & MacWhinney, 2016) and three VW experiments in German (Henry et al., 2017; Hopp, 2015; Kamide, Scheepers, & Altmann, 2003) were presented that showed successful prediction of an upcoming thematic role based on case marking information for adult L1 speakers. There is further indication that already at the age of four to five years German-speaking children can use case marking to predict a plausible object/Patient or subject/Agent (Özge et al., 2016) together with a supporting visual context (Zhang & Knoeferle, 2012). Two of the cited studies (Hopp, 2015; Mitsugi & MacWhinney, 2016) also tested L2 speakers. Both came to the conclusion that L2 speakers were unable to use this morphosyntactic cue. However, the experiment by Mitsugi and MacWhinney did not include highly proficient L2 speakers and both experiments tested L2 speakers whose L1 does not have a proper case marking system, namely English, which only marks case on pronouns and has no flexible word order. Hence, it is unclear whether a lack of L1-L2 similarity affected the outcome and the results cannot be generalized to L2 speakers as such. Other aspects that might have hindered L2 speakers from predicting concern the experimental design (object topicalization in German) and the task (switching between Japanese and English) as discussed in section 4.2. In the experimental design employed by Hopp (2015), see Table 7.2, case marking was investigated together with verb semantics. If case marking was used successfully, participants should anticipate a plausible Patient in condition (a) SVO and a plausible Agent in condition (b) OVS, thus there were two candidates in the visual scene competing with each other. The results showed that the L2 group relied on verb semantics and linear order only and anticipated the Patient of the canonical structure, in the example shown in Table 7.2 the AOI *deer*, for both linearization patterns.

			CRITICAL WINDOW	
(a) SVO	Der Wolf	tötet	gleich	den Hirsch
	The-*Nom* wolf	kills	soon	the-*Acc* deer
(b) OVS	Den Wolf	tötet	gleich	der Jäger
	The-*Acc* wolf	kills	soon	the-*Nom* hunter

Table 7.2: Experimental design in Hopp (2015). Here, case marking was investigated via object topicalization. To predict the final argument, participants had to integrate the morphosyntactic information and the lexical-semantic information given at the verb.

A design closer to the one used in Experiment 3 was employed by Mitsugi and MacWhinney (2016) in the verb-final language Japanese. Here the conditions as displayed in Table 7.3 were tested. Conditions (a) canonical and (b) scrambled were followed by a ditransitive verb and (c) accusative by a monotransitive verb. Only in (a) and (b) was a Theme argument likely to follow after the adverbial phrase, which was taken as the critical window.

			CRITICAL WINDOW		
(a) canonical	majimena gakusei-ga	kibishii sensei-ni	shizukani	tesuto-o	watashita
	serious student-*Nom*	strict teacher-*Dat*	quietly	test-*Acc*	handed over
(b) scrambled	kibishii sensei-ni	majimena gakusei-ga	shizukani	tesuto-o	watashita
	strict teacher-*Dat*	serious student-*Nom*	quietly	test-*Acc*	handed over
(c) accusative	majimena gakusei-ga	kibishii sensei-o	shizukani	karakatta	
	serious student-*Nom*	strict teacher-*Acc*	quietly	teased	

Table 7.3: Experimental design in Mitsugi and MacWhinney (2016). Here, sentences with a ditransitive verb, (a) canonical and (b) scrambled, or a monotransitive verb, (c) accusative, were tested.

A VW experiment by Kamide, Altmann, and Haywood (2003) investigated the prediction of a specific Goal based on verb semantics; see subsection 4.1.4, the example is repeated in (81). The verbs used were ditransitive and the anticipation of the plausible Goal was measured on the first post-verbal argument.

(81) a. The woman will spread the butter on the bread.
b. The woman will slide the butter to the man.

Experiment 3 combined the designs used by Mitsugi and MacWhinney (2016) and Kamide, Altmann, and Haywood (2003) and examined the use of case marking on the first post-verbal argument following a ditransitive verb to predict a plausible Theme or Recipient in German. In this experiment, the same highly proficient Russian L1-German L2 speakers as in Experiment 1 and Experiment 2 were tested. The use of case marking information was investigated by changing the word order in the German middle field. The constraints under investigation were the lexical-semantic constraint 'Recipients precede Themes' and its syntactic counterpart 'dative arguments precede accusative arguments' (Ellsiepen & Bader, 2018). Based on case marked on the first post-verbal argument (dative or accusative), participants were expected to anticipate either an accusative object/Theme or a dative object/Recipient. A simple truth-value judgment task assessed whether participants paid attention but did not specifically test their sensitivity to case information. Whether the L2 speakers were able to assign the correct case in the sentence structures they were presented with was tested in an additional offline questionnaire. To the best of my knowledge, no VW experiment in German had been conducted up to that point that had tested prediction in double object constructions. The experiment should answer the following research question:

> RQ: Do L1 and L2 speakers of German use case marking on the first post-verbal argument to predict the thematic role of the final argument?

The predictions for Experiment 3 were as follows:

1. Since previous research with L1 speakers demonstrated a predictive use of case marking (Henry et al., 2017; Hopp, 2015; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003; Mitsugi & MacWhinney, 2016; Özge et al., 2016), it was expected that the L1 group would anticipate a plausible Theme or Recipient based on the respective case marked on the first post-verbal argument. However, previous experiments in German used different sentence structures (SVO vs. OVS), where either a Patient or Agent should be anticipated based on the case marked on the NP1 and the lexical-semantic information of the verb, with the latter being a strong cue for prediction as shown in Experiment 1. In the Japanese experiment by Mitsugi and MacWhinney (2016) only one thematic role, the Theme, was expected to be anticipated, thus there was no competition between thematic roles. Since the order 'accusative object/Theme precedes dative object/Recipient' is marked, prediction was likely to be delayed compared to the canonical order, because the competitor, i.e., the Theme in the canonical condition, might be considered initially.

2. (a) If L2 speakers in general were unable to use case marking, they should show no difference in looks towards the respective target and competitor until the final argument was encountered.

(b) Another possible outcome was that the L2 speakers would over-rely on word order and, thus, show prediction for the canonical, but not the non-canonical condition. Hence, the L2 group might show the same pattern for both conditions and a prolonged preference for the Theme argument (although implausible here) in the non-canonical condition.

7.2 Methods

7.2.1 Participants, apparatus & experimental procedure

Experiment 3 was conducted together with Experiment 1 and Experiment 2 in one experimental session. The experimental items appeared within the same presentation list. For information about the participants, the apparatus and the experimental procedure, see subsection 5.2.1, subsection 5.2.3, and subsection 5.2.4, respectively.

7.2.2 Design & materials

7.2.2.1 Eye-tracking materials

Twenty-eight items in two experimental conditions, resulting in 56 sentences overall, were constructed. An example of an experimental item is given in Table 7.4. All sentences included a ditransitive verb that allows two different linearization patterns in German:

- ☐ Agent > Recipient > Theme (canonical)
- ☐ Agent > Theme > Recipient (non-canonical)

					CRITICAL WINDOW		
	NP1, Agent	verb	NP2, Recipient			adverbial	**Theme**
(a)	Der Gärtner	gibt	der	blühenden	Pflanze	eilig	frisches Wasser.
	The gardener	gives	the-*Dat*	flowering-*Dat*	plant	quickly	fresh-*Acc* water
	‘The gardener quickly gives fresh water to the flowering plant.’						
	NP1, Agent	verb	NP2, Theme			adverbial	**Recipient**
(b)	Der Gärtner	gibt	die	blühende	Pflanze	eilig	dem Postboten.
	The gardener	gives	the-*Acc*	flowering-*Acc*	plant	quickly	the-*Dat* postman
	‘The gardener quickly gives the flowering plant to the postman.’						

Table 7.4: Auditory stimulus in Experiment 3. Each item appeared in a canonical (a) and a non-canonical order (b). The critical window, shaded in grey, was cross-spliced. Case was marked on the article and adjective of the NP2 (underlined).

The article and adjective of the first post-verbal argument/NP2 mark it either as a dative or accusative object and, thus, either as a Recipient or a Theme argument. From this point onward, the final argument could be predicted, therefore the critical window started after adjective offset and at noun onset of the NP2. The critical window ended after an adverbial with a length of two to three syllables inserted between the first post-verbal argument and the final argument. The noun of the NP2 had a length of two to four syllables; together with the adverbial the critical window was four to six syllables long. The sentences were spoken by the same female native speaker of German who read the sentences for Experiment 1 and Experiment 2. Cross-splicing was applied, so that the critical window had the same length and prosody in both conditions for an item.

All nouns used as the first post-verbal argument/NP2 had either feminine or neuter gender, because dative and accusative marking for masculine gender is acoustically hard to discriminate in German and there is no difference on the adjective (Dative: *dem kleinen Jungen* vs. Accusative: *den kleinen Jungen* – the small boy). The adjective was used because it provided an additional cue for the L2 learners, whose L1 Russian has no article system but marks case on adjectives. The adjectives used could not or could hardly refer to the final argument because it was not plausible and/or it was gender incongruent. The final argument was inanimate in the canonical condition (a) and animate in the non-canonical condition (b), with only one exception.[32] The animacy (alive vs. not alive) of the first post-verbal argument varied, an aspect addressed in a post-hoc analysis.

Figure 7.1 shows the visual display for the example item in Table 7.4. The items were equally distributed across two lists using a Latin square design. The visual display for one item was always the same in both conditions, but the position of the four pictures was rotated for other items, so they would not always appear in the same position. The picture set consisted of non-colored drawings taken from the MultiPic database (Dunabeitia et al., 2018) or https://openclipart.org complemented by new drawings. All pictures of human beings had the same style.

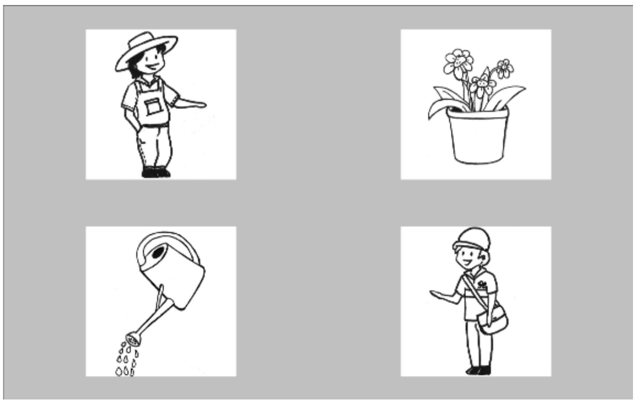


Figure 7.1: Visual display in Experiment 3. Pictures were rotated so they would not appear in the same order every time. The pictures showed the Agent (*gardener*), the first post-verbal argument (*plant*) and the Theme (*water*) and Recipient (*postman*).

Two other experimental sets (see Experiment 1 and Experiment 2) served as fillers. Actual fillers (28 in total) were also added, some mimicking the experimental manipulation in Experiment 3, see (82). This resulted in 108 sentences in total. Items and fillers appeared in a pseudorandomized order. All items and

[32]For one item the non-canonical version ended with an inanimate noun (*The policeman hands the thief directly over to the police station*).

fillers are listed in Appendix B under section B.3 and section B.4.

(82) Den schweren Ordner gibt die Sekretärin dem Forscher.
[the heavy folder]-*Acc* gives [the secretary]-*Nom* [the researcher]-*Dat*
'The secretary gives the heavy folder to the researcher.'

The L2 group was familiar with a flexible word order from their L1. Russian discriminates between six cases: nominative, accusative, genitive, dative, prepositional/locative and instrumental. The word order is only restricted to SVO or SOV in case of genuine ambiguity (Cubberley, 2002, p. 225).

7.2.2.2 Post-test

The eye-tracking materials designed were expected to follow either the canonical or a non-canonical order. However, whether the non-canonical sentences were indeed rated as less acceptable than the canonical ones had not been tested. To justify the labeling used here, the sentences were presented to a group of German native speakers, who rated their acceptability on a scale from 1 to 7, with 1 meaning 'very bad' and 7 'very good.' No fillers were included. The sentences were distributed across two lists with the same number of items per condition and randomized by-participant. This post-test was part of a web-based study conducted with Ibex farm (spellout.net/ibexfarm/). The explicit norming task followed an unrelated speeded acceptability task. After exclusion of one participant who reported having a language disorder, data from 42 participants (mean age: 29.31, range: 19–55, 24 female) were analyzed. A linear mixed-effects model as shown in Table 7.5 confirmed that the order 'Theme precedes Recipient' was rated as less acceptable than the order 'Recipient precedes Theme.' The results further indicate that the non-canonical sentences were still highly acceptable with a mean score of 5.76.

	Estimate	Std. Error	*t*-value	*p*-value
Intercept (condition = canonical)	6.3532	0.0928	68.487	< 0.001 ***
Condition	-0.5943	0.1494	-3.978	< 0.001 ***
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Formula: Response ~ condition + (1 + condition | item) + (1 + condition | subject)				

Table 7.5: Results of the materials post-test for Experiment 3

7.2.2.3 Offline test materials

To be able to anticipate a plausible candidate for the final argument, the parser first had to access the meaning of the first post-verbal argument, recognize the case marked on the article and adjective and then successfully map it to a thematic role. To assess whether the L2 learners tested in the online experiment could do so in general, an additional offline experiment was carried out. For the offline experiment, the same but slightly shortened sentences were used, as shown in Figure 7.2 below. Since the adverbials were only necessary in the online experiment to extend the critical window, they were removed for the offline experiment. The participants were instructed to choose between two completion options, so the sentence would form a grammatically correct sentence in German.

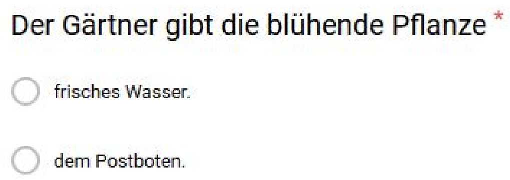


Figure 7.2: Offline questionnaire for Experiments 1–3. Participants had to indicate the grammatically correct sentence completion.

The offline test was completed after the eye-tracking experiment. Depending on the WiFi connection in the test room, the participants completed the questionnaire either on the web or received the paper version. The sentences were randomly mixed with the sentences used in Experiment 2 to also test participants' knowledge of subject-verb agreement and with fillers (see subsection 6.2.2), so there were 66 sentences in total. To ensure that participants would not simply remember which words appeared together in a sentence, they received the other list in the offline experiment, so that the final argument corresponding to the grammatically correct continuation was different.

7.2.3 Procedure: Data pre-processing

In the following, I will describe the pre-processing of the eye-tracking data. As it turned out that for five of the items in the canonical condition also the picture of the Recipient in the non-canonical condition would have been a plausible continuation, all these problematic items were excluded to make sure that only one of the pictures would show a plausible Theme argument in the canonical condition.[33] To avoid an unbalanced number of observations per word order condition, the complete items were removed. Hence, 23 item pairs were left for the analyses. As for Experiment 1 and Experiment 2, the third block for one of the L1 speakers, whose eye-tracking became very unstable, and the first block for one of the L2 speakers, who initially had problems in understanding the task, were removed. In addition, two trials had to be removed from the L1 data due to coughing and a sound problem, as well as one trial from the L2 data that was accidentally skipped plus two trials for which L2 speakers reported unknown vocabulary. Thus, of the observations left after the exclusion of the problematic items, further 1% for the L1 and 2% for the L2 group were removed.

[33]For example, after *The student presents to his current girlfriend* the Theme argument could be the inanimate object *motorbike* as intended, but also the picture of the *parents* would fit. All sentences and information about the pictures can be found in section B.3 in Appendix B. The excluded sentences are also listed there.

In Experiment 3, participants could fixate four AOIs or the background of the screen. Background looks were treated as trackloss (within the critical window 3% for L1 and 4% for L2). If case marked on the first post-verbal argument was used to predict the final argument, participants should already shift their gaze to the plausible AOI within the critical window. In the canonical condition, where the first post-verbal argument was marked as dative and thus as a Recipient, they were expected to direct their gaze to a plausible Theme. In the non-canonical condition, where the first post-verbal argument was marked as accusative and thus as a Theme, they were expected to direct their gaze to a plausible Recipient. Note, however, that the critical window for such an anticipatory effect was the time window during which participants heard or had just heard the first post-verbal argument, so it was expected to observe still more fixations on the object displaying the first post-verbal argument. The proportion of looks for all AOIs are shown in Figure 7.3 for the canonical and in Figure 7.4 for the non-canonical condition. For better interpretation, the AOIs are labeled with reference to the example item in Table 7.4, which is also shown above the graphs. The plots for all AOIs confirmed the expectations stated above: In both groups and for both experimental conditions the proportion of looks for the picture showing the first post-verbal argument (*plant*) started to increase within the critical window but, importantly, also started to decrease towards the end, while looks towards the AOI displaying the Theme for the canonical condition (*water*) and the Recipient for the non-canonical condition (*postman*), at least visibly, increased as a function of word order. The proportion of looks for the picture showing the Agent (*gardener*) clearly sank below the chance level of 0.25.

Since not of further interest, fixations on the Agent and first post-verbal argument were discarded and the relative difference between the respective target AOIs was analyzed. The exact on- and offsets determined in Praat (Boersma, 2001) were used to analyze the trajectory of the effect within the critical window. Its onset, starting at the offset of the adjective of the NP2, which was shifted 200 ms forwards, was aligned to zero. After the removal of trials with no target AOI fixations, the critical window had a mean length of 990.6 ms (SD = 137 ms). Data points were aggregated into 50 ms time bins. 81% of the L1 and 81% of the L2 On-AOI fixations within the critical time window included fixations on the Theme (*water*) or Recipient (*postman*). For all analyses and the data pre-processing the eyetrackingR package (Dink & Ferguson, 2015) was used.

(a)	Der Gärtner gibt der blühenden	Pflanze eilig	frisches Wasser
	The gardener gives [the flowering]-*Dat*	plant quickly	fresh water

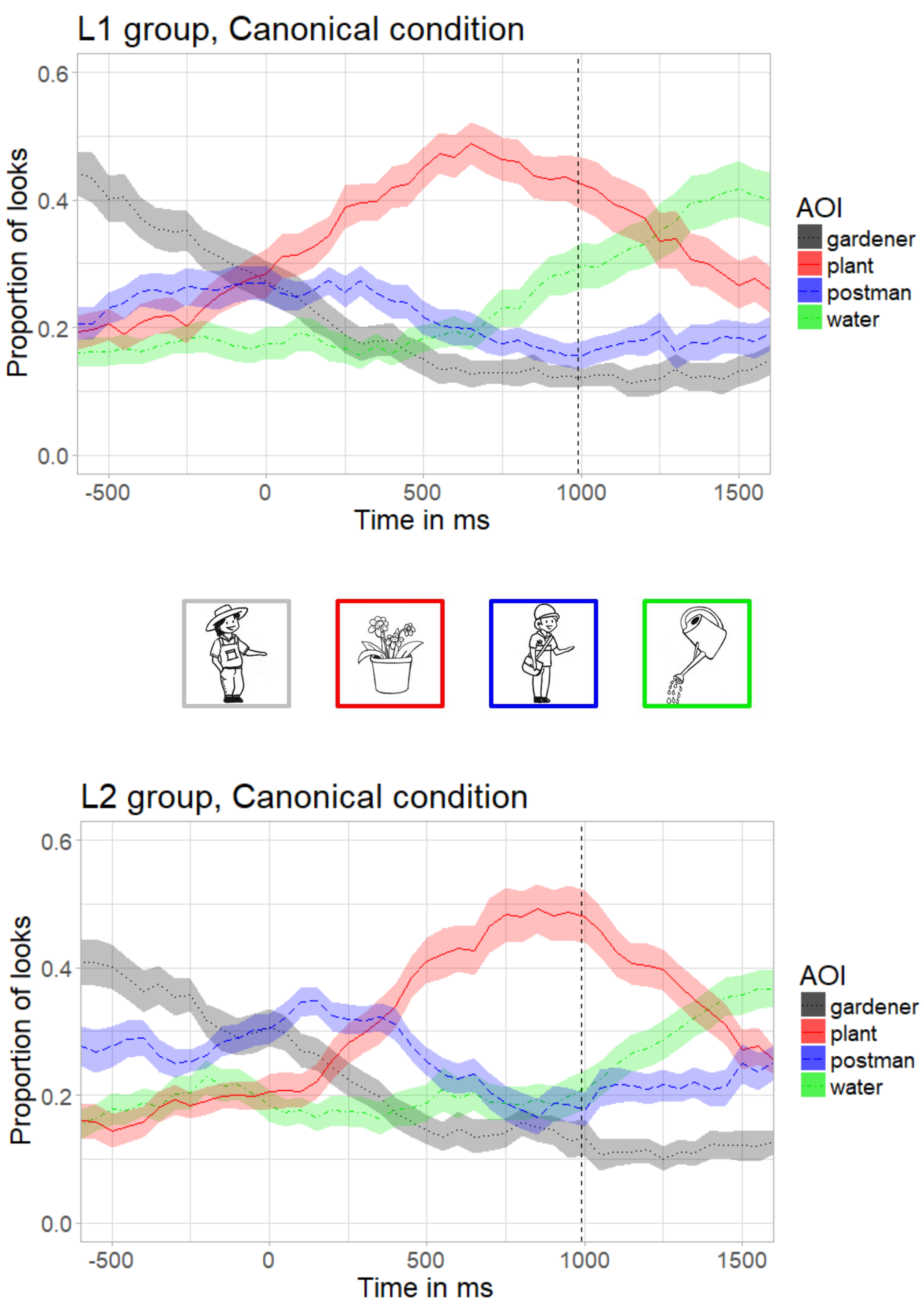


Figure 7.3: Proportion of looks for all AOIs in the canonical condition in Experiment 3 for the L1 group (top) and L2 group (bottom). The adjective offset shifted 200 ms forwards is aligned to zero, and the dashed vertical line marks the mean onset of the final argument. The sentence above together with the four pictures exemplify what the participants looked at.

(b) Der Gärtner gibt die blühende Pflanze eilig dem Postboten

The gardener gives [the flowering]-*Acc* plant quickly the postman

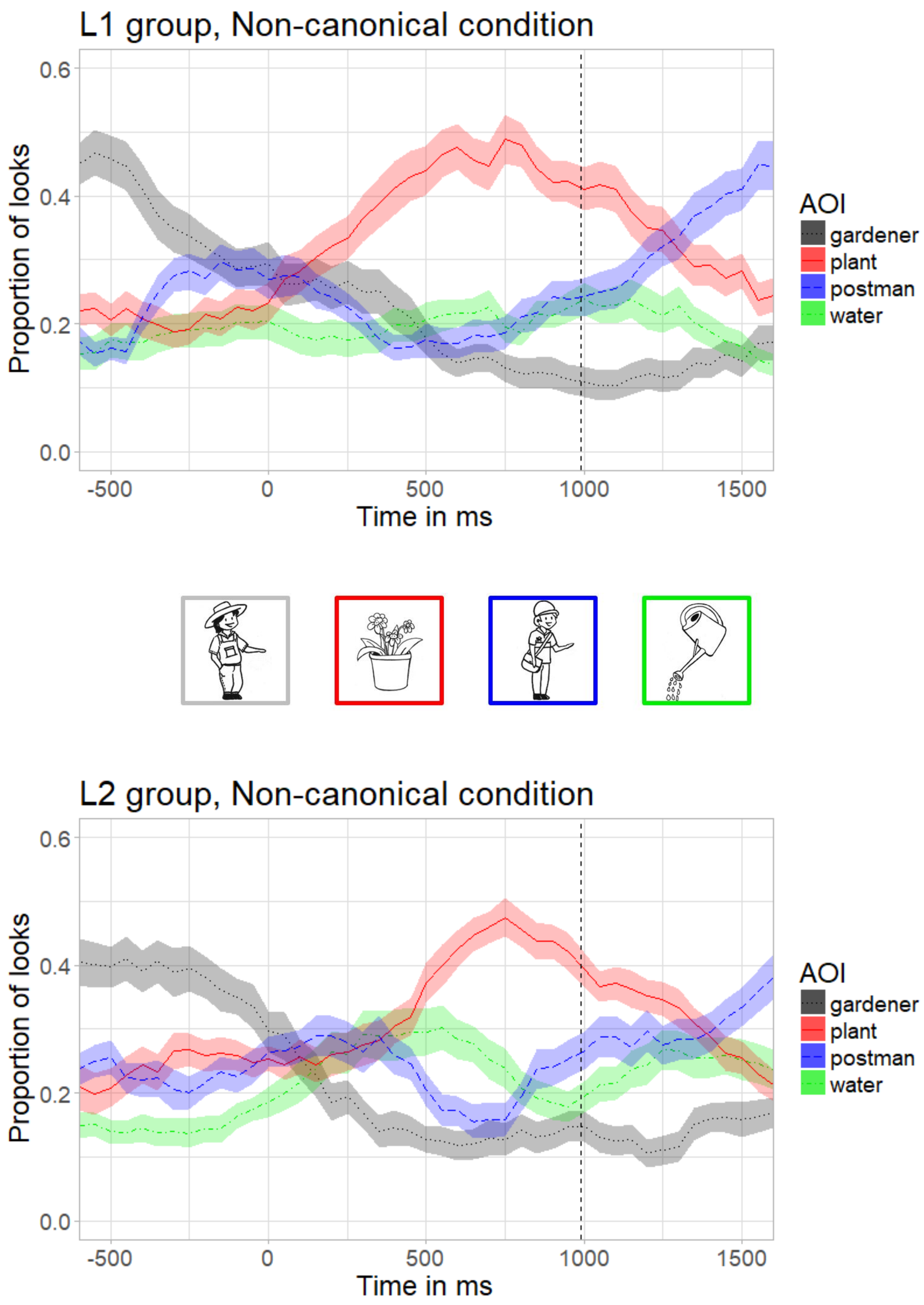


Figure 7.4: Proportion of looks for all AOIs in the non-canonical condition in Experiment 3 for the L1 group (top) and L2 group (bottom). The adjective offset shifted 200 ms forwards is aligned to zero, and the dashed vertical line marks the mean onset of the final argument. The sentence above together with the four pictures exemplify what the participants looked at.

7.3 Results

For the behavioral data, i.e., the response accuracy for the statements following half of the sentences in the eye-tracking experiment, see subsection 5.3.1.

7.3.1 Offline data

The offline test checked whether the L2 speakers were able to choose a grammatically correct continuation for a sentence context based on case marked on the first post-verbal argument. The accuracy of the L2 participants included in the analysis ranged from 82% to 100% (mean: 97%, SD: 5%), showing that all of them had sufficient knowledge of German case marking. As indicated in the participant section (subsection 5.2.1), one participant in the L2 group was excluded due to less than 80% accuracy in the offline questionnaire.

7.3.2 Eye-tracking data

For the growth curve analysis, the canonical condition was taken as the baseline to which the Elog-transformed fixation proportions for the Theme argument in the non-canonical condition were compared; see Figure 7.5 for a visualization. If case was used predictively, participants should be less likely to fixate the Theme in the non-canonical than in the canonical condition. The probability of looks to the Theme (out of total looks to the Theme and Recipient) for the canonical condition was expected to increase during the critical window. To better capture the curvature in the data, quadratic time was included as a second-order orthogonal polynomial (ot2). Note that initial looks towards the competitor, the Theme, and hence a curve were expected for the non-canonical order. The linear mixed-effects model included the factors word order condition (canonical vs. non-canonical) and group (L1 vs. L2) together with time (linear and quadratic) as fixed effects and their respective interactions. The factor group was sum contrasted (0.5, -0.5). The model with the maximal random effects structure (Barr et al., 2013) did not converge. The best fitting least maximal model as determined through model comparisons and based on the AIC value (Matuschek et al., 2017) was selected, here including subjects and items as random intercepts and the interaction between condition and group as by-item slope, as well as the interaction between condition and time (linear and quadratic) as by-subject slope; the formula is also shown below the model output in Table 7.6. The intercept in Table 7.6 corresponds to the average probability that the Theme was fixated in the canonical condition across groups. The results of a follow-up analysis are also reported in the table below the main analysis.

The model output in Table 7.6 shows no effect of word order condition across the critical window (no effect on the intercept term), but an effect of linear time, which indicates that the probability that the

Theme was fixated significantly increased for the canonical condition as predicted. In addition, the results show a significant interaction between word order condition and linear time, indicating a difference between both conditions as time increased linearly, as well as between word order condition and quadratic time. When taking the non-canonical condition as the baseline, which was done by re-leveling the factor condition, an effect of linear time was absent. However, an effect of quadratic time on the probability of looks to the Theme in the non-canonical condition in a negative direction indicates that, as time increased non-linearly, participants were less likely to fixate the Theme. There was no significant effect of group or interaction with group, indicating that any differences between the L1 and L2 groups did not approach significance. All effects and interactions observed were across groups.[34]

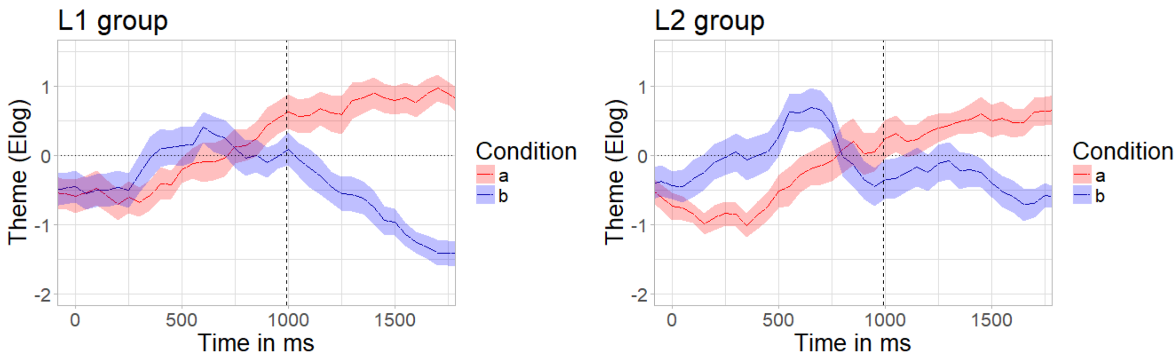


Figure 7.5: Time course showing the probability of looks to the Theme for the canonical (a) and non-canonical (b) conditions in Experiment 3 for the L1 group (left) and L2 group (right), starting at adjective offset shifted 200 ms forwards. The dashed vertical line marks the mean onset of the final argument.

To sum up, the probability that the Theme was fixated in the canonical condition significantly increased (linearly), demonstrating beginning anticipation of the target argument. This effect was observed across groups. As can be seen in the graphs in Figure 7.5, both groups also initially considered the Theme argument (although implausible) in the non-canonical condition. In the inferential statistics, this was reflected in terms of a missing effect of condition and no effect of linear time in a negative direction for the non-canonical condition, which would indicate that the probability that the Theme was fixated decreased. However, the probability of looks to the Theme did not increase with the same slope as it increased for the canonical condition as indicated by the interaction between condition and time. Instead, there was an effect of quadratic time on the probability of looks to the Theme in a negative direction for the non-canonical condition: As time increased (non-linearly), the probability that the Theme was fixated decreased. Hence, the anticipatory effect developed differently for the two word order conditions.

[34]The results of the omnibus model with treatment contrast coding for the factor group are shown in Table B.6 in Appendix B. The effect of linear time for the canonical condition was significant in both groups. Both groups also showed a marginally significant interaction between condition and time, indicating a different development for the probability of looks to the Theme for the canonical and non-canonical conditions.

	Estimate	Std. Error	*t*-value	*p*-value
Intercept (condition = canonical)	-0.0762	0.1791	-0.426	0.6724
Condition (non-canonical)	0.0674	0.2216	0.304	0.7621
Group	0.2908	0.295	0.986	0.3286
Linear time	2.8112	0.5977	4.703	< 0.001 ***
Quadratic time	0.6029	0.4456	1.353	0.1821
Condition x Group	-0.4831	0.4157	-1.162	0.2498
Condition x Linear time	**-2.0369**	**0.8502**	**-2.396**	**0.0206 ***
Condition x Quadratic time	**-1.5676**	**0.6273**	**-2.499**	**0.0158 ***
Group x Linear time	0.2122	1.195	0.178	0.8598
Group x Quadratic time	-0.1418	0.891	-0.159	0.8742
Condition x Group x Linear time	-0.4047	1.6998	-0.238	0.8128
Condition x Group x Quadratic time	0.3766	1.2543	0.300	0.7652
Intercept (condition = non-canonical)	-0.0088	0.1759	-0.050	0.9602
Linear time	0.7743	0.6602	1.173	0.2469
Quadratic time	-0.9647	0.3939	-2.449	0.0183 *

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Formula: Elog ~ condition*group*(ot1+ot2) + (1+condition*group | item) + (1+condition*(ot1+ot2) | subject)

Table 7.6: Results of the growth curve analysis for the critical window in Experiment 3. Below the output of the model with the canonical condition as baseline, the results of the follow-up analysis based on the interactions (highlighted) are shown.

To test whether the emerging anticipatory effect seen in the analysis of the critical window actually led to an overall effect of word order in the post-critical window, the 650 ms time window after the onset of the final argument (shifted 200 ms forwards) was further analyzed. The final argument differed between items insofar as it sometimes included an article and adjective or only an adjective followed by a noun. A fixed length was chosen here based on the average duration of the final NP, which had a mean length of 655.1 ms (SD = 172.93 ms). Again, the maximal model did not converge and the best fitting least maximal model that converged was selected based on model comparisons. Due to interactions with the factor group, the same linear mixed-effects model with treatment contrast coding for group was computed, the results of which are shown in Table 7.7. First, the results of the omnibus model are shown with the L1 group as reference group and, after re-leveling of the factor group, with the L2 group.

The effect at the intercept for the L1 group shows that the likelihood that the Theme was fixated in the canonical baseline condition rose significantly above chance. The effect of condition shows that the L1 group was less likely to fixate the Theme in the non-canonical condition. The beginning anticipatory effect thus resulted in an overall effect in the post-critical window. The interaction between word order condition and linear time for the L1 group further indicates a continuing decrease in the likelihood that

	Estimate	Std. Error	*t*-value	*p*-value
Intercept (condition = canonical, L1 group)	0.7683	0.2541	3.023	0.0041 **
Condition (non-canonical)	-1.2544	0.33287	-3.768	0.0004 ***
Group (L2)	-0.5744	0.31783	-1.807	0.0762 .
Linear time	0.3904	0.3253	1.200	0.2362
Quadratic time	-0.2221	0.1652	-1.345	0.1788
Condition (non-canonical) x Group (L2)	0.7694	0.4601	1.672	0.0999 .
Condition x Linear time	-1.5833	0.51497	-3.075	0.0035 **
Condition x Quadratic time	-0.2518	0.2301	-1.094	0.274
Group x Linear time	-0.4692	0.47530	-0.987	0.3286
Group x Quadratic time	0.6172	0.24248	2.545	0.0109 *
Condition x Group x Linear time	1.4085	0.7518	1.874	0.067 .
Condition x Group x Quadratic time	0.042	0.3396	0.124	0.9015
Intercept (condition = canonical, L2 group)	0.1939	0.2642	0.734	0.4666
Condition (non-canonical)	-0.485	0.3537	-1.371	0.1762
Linear time	-0.0788	0.3466	-0.227	0.8210
Quadratic time	0.3951	0.1775	2.226	0.0261 *
Condition x Linear time	-0.1748	0.5477	-0.319	0.751
Condition x Quadratic time	-0.2098	0.2498	-0.840	0.401

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Formula: Elog ~ condition*group*(ot1 + ot2) + (1 + condition*group | item) + (1 + condition*(ot1) | subject)

Table 7.7: Results of the growth curve analyses for the 650 ms post-critical window in Experiment 3. The results show the output of the omnibus model with the L1 group as reference group (top) and the L2 group (bottom).

the Theme was fixated in the non-canonical condition. When taking the non-canonical condition as the baseline by re-leveling the factor condition, there was an effect of linear time (Est. = -1.193, SE = 0.308, t = -3.877, p = 0.0003), as well as an effect of quadratic time (Est. = -0.4739, SE = 0.1601, t = -2.960, p = 0.0031) in a negative direction. This lasting effect indicates that the competition between Theme and Recipient in the non-canonical condition spilled over into the post-critical window, so the effect of word order condition further developed over time. In contrast, in the L2 group no overall effect of condition emerged, at least not within the first 650 ms of the post-critical window, but fixations reached some kind of plateau. This difference led to an effect of group, although only as a trend. An effect of quadratic time for the canonical condition indicates that the probability that the Theme was fixated continued to increase in a non-linear fashion. This difference between groups further shows up as an interaction between group and quadratic time. In addition, a marginal three-way interaction between word order condition, group and linear time indicates that the linear decrease in the probability that the Theme was fixated in the non-canonical condition tended to be stronger in the L1 group.

To sum up the results of the analyses of the post-critical window, L2 speakers' information integration was delayed for the canonical condition and more affected by the word order variation in the non-canonical condition. Although the L2 group showed a similar pattern for the critical window, the probability that the Theme was fixated in the canonical condition further increased gradually over time. Hence, whereas anticipation led to an immediate preference for the respective target in the L1 group, the L2 group appeared to be slower or less certain, especially in the non-canonical condition.

When creating the materials for this experiment, I looked for entities as first post-verbal arguments that could either receive something, thus functioning as a Recipient, or be given to someone, thus functioning as a Theme. However, there is another constraint that may play a role here, namely animacy. As mentioned in the design and materials section, the final argument in the canonical condition (a) with the order 'Recipient precedes Theme' was always inanimate and in the non-canonical condition (b) with the argument order 'Theme precedes Recipient' it was always animate with only one exception. Although all entities as first post-verbal arguments were either animate or somehow animated, they differed in terms of where they should be placed along the animacy scale (see also section 5.1). Instead of only discriminating between animate and inanimate, the distinction can be even more gradient with humans ranked above animals, and then it can go even further by taking into account concepts like locomotion or sentiency, which also includes intentionality (Yamamoto, 2006, p. 29–37).

In a post-hoc analysis, such a discrimination was applied to the items in Experiment 3. All items where the first post-verbal argument was either a human being or an animal were categorized as "animate" (13 items) and all objects and plants were pooled together as "inanimate" (10 items). Note that plants ideally should be considered as animate, however compared to humans and animals they appear less animate and closer to the inanimate category, which includes immobile objects, in regard to the concepts of locomotion and sentiency mentioned before. Therefore, and also for the sake of better balancing the number of items, plants (2 items) were treated as inanimate here. The last column in Table 7.8 presents the ranking of these new, more fine-grained conditions.

Altogether word order condition (a) should still be easier to process, because it follows the canonical order 'dative object/Recipient precedes accusative object/Theme' and, hence, these conditions are marked with a plus. However, since Recipients are typically animate and animate objects are placed in front of inanimate objects, word order condition (a) together with an animate Recipient received an additional plus. The two animacy orders in condition (b) should not vary; both are non-canonical and the constraint that animate objects precede inanimate objects did not apply. However, since Themes are typically inanimate, the order inanimate precedes animate might facilitate the processing of the non-canonical

word order	animacy	1st post-verbal argument	2nd post-verbal argument	constraints in alignment
		Recipient, dative	**Theme, accusative**	
a	a	animate	inanimate	+ +
a	i	inanimate	inanimate	+
		Theme, accusative	**Recipient, dative**	
b	a	animate	animate	- -
b	i	inanimate	animate	-

Table 7.8: More fine-grained distinction of the experimental conditions in Experiment 3. The last column indicates which conditions should be more preferred (+) or dispreferred (-).

order. Therefore, the non-canonical orders received a minus, and the condition where the Theme was animate and, thus, more likely to be mistakenly taken for a Recipient was given an additional minus.

To summarize, if animacy as a lexical-semantic constraint had an additional impact, it should exert an effect for the canonical condition in those sentence contexts in which the first post-verbal argument was animate. Here, prediction might be faster/easier than when it was inanimate. An example of a sentence in which all constraints were in alignment is shown in (83). In addition, the initial competition between Theme and Recipient in the non-canonical condition, seen in the previous analyses, might be reduced when the thematic role was not only morphosyntactically encoded, but also had the prototypical animacy feature.

(83) Der Vater überreicht dem schreienden Baby vorsichtig den Schnuller.
[The father]-*Nom* hands [the crying baby]-*Dat* carefully [the pacifier]-*Acc*
'The father carefully hands the pacifier to the baby.'

The graphs in Figure 7.6 show the two word order conditions, canonical (a) and non-canonical (b), for each group displayed next to each other, further split into the two animacy conditions, so there were four levels overall. *Condition aa versus ai:* The red line showing the probability of looks to the Theme for the condition where the first post-verbal argument was animate rose above the blue line for the inanimate condition when the word order followed the canonical order 'dative object/Recipient precedes accusative object/Theme.' Hence, when all constraints were in alignment as in (83), the increase appeared to be steeper, not only for the L1 but also the L2 group. *Condition ba versus bi:* For the non-canonical word order, the probability that the Theme was fixated was slightly reduced for the condition where the first post-verbal argument was inanimate, as indicated by the blue line (although note the spike towards the end of the critical window in the L1 group). Interestingly, there is a visible difference between groups after the onset of the final argument. The L2 group showed once again a rising probability of looks to the implausible inanimate Theme after an animate Theme argument, as indicated by the red line. At least visibly, the two ad-hoc hypotheses proposed above seem confirmed.

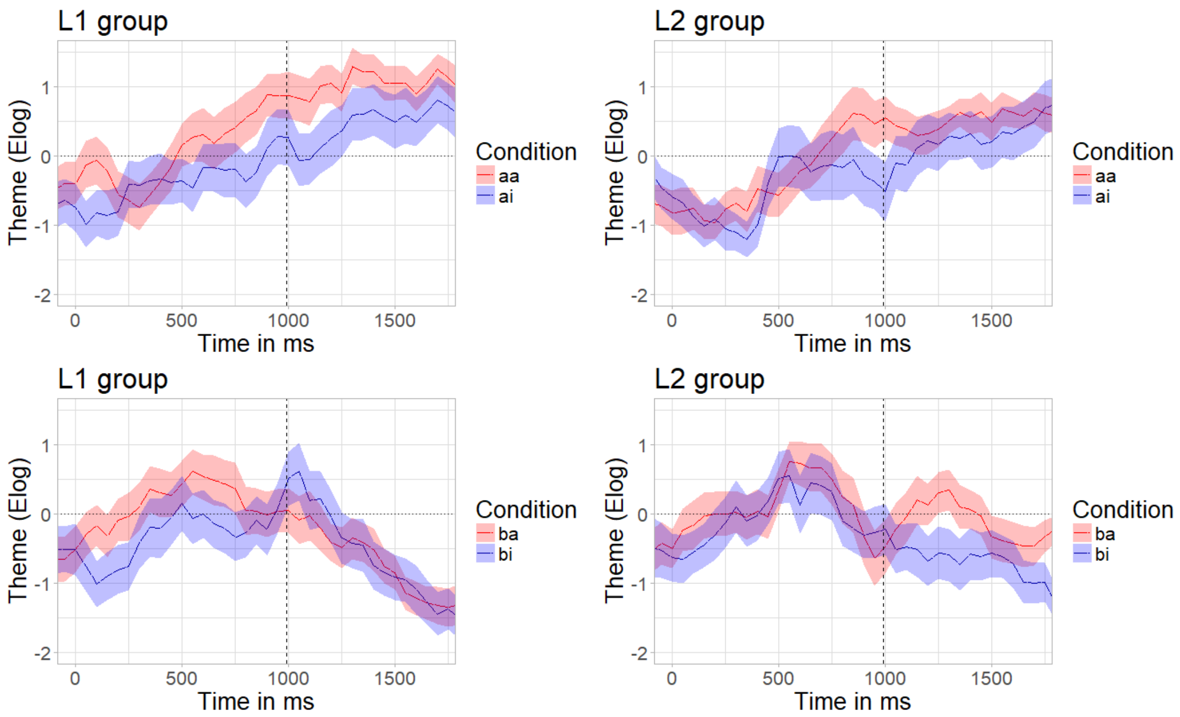


Figure 7.6: Time course per condition showing the probability of looks to the Theme, starting at adjective offset shifted 200 ms forwards. The canonical condition (a) displayed at the top and non-canonical condition (b) in Experiment 3 are further split into a condition where the first post-verbal argument is animate, (aa) and (ba), or inanimate, (ai) and (bi). The dashed vertical line marks the mean onset of the final argument. The plots on the left show the time course for the L1 group, the plots on the right the time course for the L2 group.

The new conditions were entered into a linear mixed-effects model. In Table 7.9, the output of the best-fitting model that converged is shown. The fixed effects included the now between-items factor condition with four levels (canonical, animate (aa) vs. canonical, inanimate (ai) vs. non-canonical, animate (ba) vs. non-canonical, inanimate (bi)), the factor group (sum contrasted) and time (linear and quadratic), as well as their respective interactions. The random effects included subjects and items as random intercepts, as well as the interaction between group and quadratic time as by-item slope, and condition as by-subject slope. To reduce complexity, the interactions between conditions and group, and group and time, as well as their respective three-way interactions in Table 7.9 are dropped, none of which approached significance. To get the estimates in comparison to different reference levels, the omnibus model was taken and the factor condition re-leveled.

The results show a main effect of linear time on looks to the Theme across groups for both canonical conditions, independent of the animacy of the first post-verbal argument. Furthermore, for both canonical conditions, when taken as the reference level, an interaction between the non-canonical conditions and linear time emerged, which indicates that the linear increase was more pronounced for the canonical orders. A further interaction between the non-canonical conditions and quadratic time, when condition (aa) was taken as the reference level, points to a difference in the non-linear development. For both canonical conditions, the estimates for quadratic time showed a positive development and for the non-canonical conditions a negative development, meaning a decrease over time in the probability that the Theme was fixated. However, only for condition (bi) was there a significant effect of quadratic time across groups, indicating a non-linear decrease for the probability that the Theme was fixated. When condition (ai) was taken as the reference level, only the interaction between condition (bi) and quadratic time reached significance. There was also a linear increase in the probability of looks to the Theme for the non-canonical conditions, which, however, was only marginal for condition (bi), where the Theme argument was inanimate. Hence, as hypothesized above, the competition between the Theme and Recipient seemed reduced if the accusative object also had the prototypical semantic feature of a Theme. However, this conclusion has to be treated with caution as there was no significant difference between the non-canonical orders. Although the increase in the probability that the Theme was fixated appeared steeper for condition (aa), where the word order was not only canonical but the dative object/Recipient was also animate, there was no significant difference between the canonical orders. Hence, the post-hoc analysis points to a modulating effect of animacy that, however, only shows up in interactions and not in within-word order comparisons. Further note that the post-hoc analysis included fewer data points than ideally needed to draw a firm conclusion regarding the effect of animacy.

Condition (aa) as reference level	Estimate	Std. Error	*t*-value	*p*-value
Intercept (L1/L2)	-0.0103	0.2027	-0.051	0.9598
Condition (ai)	-0.1843	0.3021	-0.610	0.5448
Condition (ba)	0.1301	0.1769	0.736	0.465
Condition (bi)	-0.1716	0.318	-0.539	0.5919
Group	0.4217	0.3371	1.251	0.2155
Linear time	2.7944	0.3534	7.907	<0.0001 ***
Quadratic time	0.5584	0.4891	1.142	0.2613
Condition (ai) x Linear time	-0.6825	0.5279	-1.293	0.1961
Condition (ba) x Linear time	-1.9999	0.4688	-4.266	<0.0001 ***
Condition (bi) x Linear time	-2.1159	0.52	-4.069	<0.0001 ***
Condition (ai) x Quadratic time	-0.1922	0.7407	-0.259	0.7968
Condition (ba) x Quadratic time	-1.1148	0.3938	-2.831	0.0047 **
Condition (bi) x Quadratic time	-1.8422	0.7415	-2.485	0.0178 *
Condition (ai) as reference level	Estimate	Std. Error	*t*-value	*p*-value
Intercept (L1/L2)	-0.1946	0.2255	-0.863	0.3927
Condition (ba)	0.3144	0.3144	1.000	0.322
Condition (bi)	0.0127	0.2223	0.057	0.9545
Group	0.1322	0.3708	0.357	0.7226
Linear time	2.1119	0.3922	5.385	<0.0001 ***
Quadratic time	0.3662	0.5563	0.658	0.5147
Condition (ba) x Linear time	-1.3174	0.5203	-2.532	0.0114 *
Condition (bi) x Linear time	-1.4333	0.5356	-2.676	0.0075 **
Condition (ba) x Quadratic time	-0.9227	0.7388	-1.249	0.22
Condition (bi) x Quadratic time	-1.6501	0.4642	-3.555	0.0004 ***
Condition (ba) as reference level	Estimate	Std. Error	*t*-value	*p*-value
Intercept (L1/L2)	0.1198	0.2077	0.577	0.5664
Condition (bi)	-0.3016	0.3127	-0.965	0.3392
Group	-0.1379	0.349	-0.395	0.694
Linear time	0.7945	0.342	2.323	0.0202 *
Quadratic time	-0.5565	0.4861	-1.145	0.2602
Condition (bi) x Linear time	-0.116	0.5123	-0.226	0.8209
Condition (bi) x Quadratic time	-0.7274	0.7395	-0.984	0.3321
Condition (bi) as reference level	Estimate	Std. Error	*t*-value	*p*-value
Intercept (L1/L2)	-0.1818	0.24813	-0.733	0.4668
Group	-0.2882	0.4247	-0.678	0.4999
Linear time	0.6786	0.3815	1.779	0.0753 .
Quadratic time	-1.2839	0.5573	-2.304	0.0272 *

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Formula: Elog ~ condition*group*(ot1+ot2) + (1 + group*(ot2) | item) + (1 + condition | subject)

Table 7.9: Results of the post-hoc analyses for the critical window in Experiment 3

7.4 Discussion

Experiment 3 examined whether L1 and L2 speakers of German, after encountering a ditransitive verb, make use of case marked on the first post-verbal argument to anticipate the final argument, a Theme after a dative object/Recipient and a Recipient after an accusative object/Theme. Thus, to be able to predict the final argument, participants had to access the meaning of the first post-verbal argument, recognize the case marking (dative or accusative), map it to a thematic role (Recipient or Theme) and exploit this information to predict the upcoming thematic role (Theme or Recipient). The visual display showed four pictures, the subject/Agent, the first post-verbal argument and the target object as well as the respective competitor, meaning that the plausible Theme for the canonical condition was also displayed in the non-canonical condition and functioned here as the competitor.

For the analyses, the canonical word order condition (Dat > Acc) was taken as the baseline and looks to the Theme argument (out of looks to the Theme and Recipient) were analyzed dependent on word order condition and group. A predictive use of case marking, thus, should be reflected in an increasing probability of looks to the Theme for the canonical condition and, over time, fewer looks to the Theme in the non-canonical (Acc > Dat) relative to the canonical condition. The results from Experiment 3 showed that both the L1 and the L2 groups anticipated the plausible Theme argument for the canonical condition: Before the final argument was encountered, the likelihood that the Theme was fixated increased in a linear fashion. Importantly, the same increase was absent for the non-canonical condition. Instead, over time participants across groups were less likely to the fixate the Theme argument in the non-canonical condition. Hence, unlike in the experiment by Hopp (2015), the results showed a difference between word order conditions (here: Dat > Acc vs. Acc > Dat) across groups, signaling sensitivity to case marking for both. Moreover, the fact that there was a difference between word order conditions in how the likelihood that the Theme was fixated developed over time demonstrates that the L2 group did not over-rely on word order to predict. Otherwise the same linear increase in the probability of looks to the Theme should have been observed for both word order conditions.

As predicted, the L1 but also the L2 group showed an emerging preference for the Theme over the Recipient in the canonical condition, as indicated by an effect of linear time. Moreover, a decrease in the probability that the Theme was fixated for the non-canonical condition led to an interaction between word order condition and both time terms across groups. These effects showed up in the critical window, before the final argument was encountered, and, for the L1 group, turned into an overall effect of condition upon its auditory presentation, i.e., an effect at the intercept and an effect of word order condition, indicating that the probability of looks to the Theme rose above chance level for the canonical condition and there was a significant difference between both conditions. The conclusion that the difference between conditions that developed over time in the critical window indicates an emerging anticipatory effect hence appears legitimate. A continuing decrease in the probability that the Theme was fixated in the L1 group for the non-canonical condition after the onset of the final argument pointed to initial competition between the Theme and Recipient, with a lasting effect in the post-critical window. Hence, the non-canonical order led to initial competition between the target and competitor, which were both shown in the visual display. As further discussed below, this finding is in line with recent assumptions that effects in reaction to non-canonical orders may reflect a prediction mismatch.

Although there were no between-group differences for the critical window, differences showed up in the 650 ms post-critical window. Unlike the L1 group, the L2 group showed no overall effect of condition for the post-critical window. Instead, the probability that the Theme was fixated in the canonical condition further increased gradually, potentially signaling more uncertainty in the L2 group. Hence, they did not benefit from prediction in the same way as the L1 speakers. Like the L1 speakers, they showed the same, but at least visually slightly more pronounced, initial competition between Theme and Recipient in the critical window, which, however, seemed to have a more severe effect later on: At least initially, there was no further decrease in the probability of looks to the Theme for the non-canonical condition. Future research may investigate more systematically which effect the animacy of arguments has, i.e., whether prediction and later integration of an argument is facilitated in the case of prototypical animacy features. The findings here suggest that L2 speakers were less certain than L1 speakers and needed more time to re-analyze.

To get back to the research question addressed, the results from Experiment 3 showed that L1 and L2 speakers of German successfully mapped the case marked on the first post-verbal argument to a thematic role and used this information to anticipate a plausible accusative object/Theme in the canonical condition. Moreover, both groups became aware that the Theme was not a plausible argument in the non-canonical condition. As far as the critical window is concerned, Experiment 3 found no differences between L1 and L2 speakers, with both groups showing a similar pattern. The results showed no statistically backed-up over-reliance on word order for prediction, i.e., not the same effect of linear time on looks to the Theme in the non-canonical as in the canonical condition. Hence, not only L1 but also L2 speakers of German can use case marking on the first post-verbal argument to predict the thematic role of the final argument, although most clearly for the canonical order.

In the following, I will discuss potential reasons why, in the current experiment, in contrast to previous experiments, a predictive use of case was also found for a group of L2 speakers and significant differences between groups only emerged later. I will further discuss the implications the findings have for the online processing of double object constructions and the interaction of cues.

The current experiment differed from previous experiments in several regards: First, a group of L2 speakers who have an L1 with a rich morphological case marking system was tested. Like German, Russian also allows word order variation and grammatical/thematic roles are morphologically encoded, unlike English, the L1 of the L2 groups in previous experiments, which relies on linear order. Differently or at least to a greater extent than in German, word order variation in Russian is sensitive to the discourse, however here the sentences were presented in isolation (null contexts). In a paper by Bornkessel et al. (2005) referred to in the introduction, the authors speculate that there may be cross-linguistic differences in the specialization of components in the brain. They differentiate two regions that are differently engaged in argument hierarchization depending on morphosyntactic and positional information. In subsection 3.2.2, the potential role of language similarity was reported as one factor that can influence predictive processing. However, to be able to draw a safe conclusion in this regard, an additional group of L2 speakers needs to be tested, for example English-German learners, who have the same proficiency level and same amount of exposure as the L2 speakers tested here, but whose L1 lacks a proper case marking system. Second, although the L2 speakers in Mitsugi and MacWhinney (2016) demonstrated knowledge of Japanese case marking offline, they were not highly proficient. It is further unclear what impact the switching of languages between stimulus sentences and comprehension questions had for them. Here, the L2 speakers fell in the category "advanced to near-native" as defined by Hopp (2015), who tested German SVO compared to OVS structures. I argue that it is likely that the non-canonical structure used here was less marked, so it differed from the non-canonical structure OVS in the ease of processing.[35] Note that object topicalization is further intertwined with information structure and, thus, involves the external discourse interface, which might be even more difficult for L2 speakers. Although the L2 speakers (but also the L1 speakers) showed no clear evidence for anticipation of the dative object/Recipient here, they became aware that the Theme was not a plausible final argument. The post-test for the materials showed that the non-canonical word order was rated as less acceptable by German native speakers, but that it was still highly acceptable. In addition, and related to the first aspect, language similarity, case was marked on articles, which do not exist in Russian, as well as on adjectives. Moreover, critical case marking was acoustically more pronounced as no masculine gender nouns were used, where the difference between accusative and dative is only marked on the article via *-n* versus *-m*.

Another difference regarding the materials used here and those by Hopp (2015) is that in his materials the lexical-semantics of the verb provided a further cue in addition to the case marked on the NP1. Hence, here participants had to integrate different information sources. It is possible that the English L1-German L2 speakers showed anticipatory eye-movements towards the depicted *deer* after the accusative marked NP1 *the wolf* followed by the verb *kill* because the word order/positional cue and the semantic cue received more weight than the morphosyntactic cue when sources of information had to be integrated to form a prediction.

[35]Note though that after the experiment, some Russian participants reported that these sentences sounded unusual.

Another interesting finding was that L2 and L1 speakers were affected by competition between the Theme and Recipient in the non-canonical condition, but that the L1 speakers rapidly recovered from it. Although designed to test for effects of prediction, this finding from Experiment 3 indicates that not only prediction but also re-analysis can differ between L1 and L2 speakers (e.g., Pozzan & Trueswell, 2016), which contrasts with what has been claimed by Phillips and Ehrenhofer (2015) (see subsection 3.1.3). One may further assume additional processing demands for the non-canonical condition, although it has to be pointed out that the visual-world eye-tracking paradigm is not the method of choice to test for processing load. However, the finding is in line with the assumption that previous ERP findings showing a negativity after non-canonical structures reflect a prediction mismatch (Dröge et al., 2016). Whereas Mitsugi and MacWhinney (2016) reported no difference between the canonical and scrambled condition for Japanese native speakers in accordance with previous findings on Japanese, Experiment 3 found a difference between the canonical condition and the condition that disrespected the two constraints 'dative arguments precede accusative arguments' and 'Recipients precede Themes,' which fell together in the current design. However, the two designs differed insofar as Mitsugi and MacWhinney just switched the dative and nominative marked arguments (see Table 7.3), whereas here the first post-verbal argument remained the same, but received either dative or accusative case. Since the visual display was the same for both word order conditions (see Figure 7.1), there were two potential candidates for the final argument.

A post-hoc analysis showed that animacy is a potential modulating factor and it seems worthwile to further explore its role in the processing of double object constructions more systematically. Here, the results still showed an overall effect of word order condition, with a more linear increase in the probability of looks to the Theme for the order 'dative object/Recipient precedes accusative object/Theme' and a non-linear decrease for the non-canonical order 'accusative object/Theme precedes dative object/Recipient,' which, however, was only significant in the case of inanimate Themes as the first post-verbal argument. To further examine the impact of different cues like case marking, word order and especially animacy and get a deeper understanding of the online processing of double object constructions, it would be further interesting to include a group of children, particularly in comparison to L2 learners. Constructions including two objects seem to pose a challenge for children at a certain point in acquisition. Drenhaus and Féry (2008) report that German children aged 3;9 to 6;8 years had difficulties in repeating double object (full NP) structures in an elicited imitation task when the direct/accusative object preceded the indirect/dative object, see (84), and reversed the order in 54% of the repetitions. Independent of word order, the Benefactor (here I use the term Recipient) was more often marked as accusative, as was the case for the majority of utterances, when the Theme was inanimate. Also, in sentences where the Theme was a pronoun and should be placed before the full NP, children tended to use the accusative on the Benefactor when the Theme was inanimate (85-a) and, thus, used a double accusative construction. Following the authors' explanation, the children may have felt no need to disambiguate the thematic roles of the arguments using case marking. The authors further concluded that children chose performance strategies in terms of markedness and case hierarchy that were economical, but clearly not adultlike. The children showed a preference for unmarked/canonical over non-canonical structures and omitted case marking when the lexical-semantic constraint 'animate > inanimate' was respected.

(84) a. Der Mann will das Auto dem Kind geben.
the man wants [the car]-*Acc* [the child]-*Dat* give
'The man wants to give the car to the child.'

b. Der Mann will die Katze dem Kind geben.
the man wants [the cat]-*Acc* [the child]-*Dat* give
'The man wants to give the cat to the child.'

(85) a. Der Mann will ihn [= den Stuhl] der Frau geben.
the man wants it-*Acc* [the woman]-*Dat* give
'The man wants to give it [= the chair] to the woman.'

b. Der Mann will ihn [= den Hund] der Frau geben.
the man wants it-*Acc* [the woman]-*Dat* give
'The man wants to give it [= the dog] to the woman.'

The authors state that in adult German animacy does not interact with case assignment, whereas constraints are differently ranked in children. Once acquired, case may become more economical. They further mention that "one could speculate that the preference for placing animate arguments in front of inanimate ones in the adult language might even be a relic of an ontologically older stage of grammar where this feature played an active role" (p. 238).

Altogether, Experiment 3 extended the investigation of the predictive use of case marking to double object constructions. Since not tested before, it was explorative and there are several questions not yet answered. Moreover, it opens up several new questions, including but not limited to the questions of how further L2 groups and children process these constructions, how cues interact and whether the processing of non-canonical structures would be different using an experimental design without competition between two visually shown candidates, as in example (86), which is parallel to the manipulation in Mitsugi and MacWhinney (2016). In the examples below the critical window is highlighted in bold. Here the target picture (baby) is the same for the canonical (86-a) and the non-canonical order (86-b).

(86) Miriam sieht, dass … ('Miriam sees that … ')

a. der überforderte Vater der besorgten Mutter **vorsichtig** das schreiende
[the overwhelmed father]-*Nom* [the concerned mother]-*Dat* carefully [the crying
Baby gibt.
baby]-*Acc* gives
'the overwhelmed father gives the crying baby to the concerned mother'

b. der besorgten Mutter der überforderte Vater **vorsichtig** das schreiende
[the concerned mother]-*Dat* [the overwhelmed father]-*Nom* carefully [the crying
Baby gibt.
baby]-*Acc* gives

7.5 Conclusion

Experiment 3 tested L1 and L2 speakers of German on their ability to use case to predict an upcoming argument after ditransitive verbs in double object constructions. The results showed that for a canonical structure like *Der Gärtner gibt der blühenden Pflanze eilig* (The gardener-*Nom* gives the flowering plant-*Dat* quickly) the probability of looks to the target accusative object/Theme *water* increased linearly across groups. Importantly, a different development was observed for the non-canonical structure *Der Gärtner gibt die blühende Pflanze eilig* (The gardener-*Nom* gives the flowering plant-*Acc* quickly), where a plausible Recipient should be predicted. Again, this difference between conditions was found across groups. Experiment 3, for the first time, demonstrated that L2 speakers were able to map the case marked on an argument to a thematic role in sufficient time to show a difference in the time course between two word order conditions before the onset of the final argument. Hence, when familiar with case from their L1 and given all chances to use case predictively, here case was marked on the article and adjective, L2 speakers could do so. However, although there were no between-group differences for the anticipation window, the L2 group needed more time to integrate incoming information and was affected by the non-canonical word order to a greater extent than the L1 group. The last finding points to an over-reliance on word order, a surface-level cue, although here the effect spilled over into the post-critical window. The results are in line with the assumption that L2 speakers show a different weighting of cues and rely more on surface-level and less on grammatical information than L1 speakers during real-time processing. However, the results from Experiment 3 are not in line with the assumptions of the RAGE hypothesis but rather suggest that L2 speakers also differ from L1 speakers in reactive processing.

Chapter 8

Experiment 4: Prediction based on implicit causality information

8.1 Introduction

The following experiment focuses on prediction at the discourse-level, the prediction of upcoming (co)reference. According to the IH, phenomena involving the external discourse interface should be problematic even for highly proficient L2 speakers. As a possible explanation, Sorace (2011) mentions that the allocation of cognitive resources may be different between L2 and L1 speakers. If cognitive factors play a role, one would also expect differences between child L1 and adult L1 processing. To test for this possibility, Experiment 4 also included a group of children. The predictive cue in this experiment was the implicit causality bias of psych verbs mentioned in subsection 4.1.6 and subsection 4.2.6. Here, the effect of the predictive cue was investigated by looking at temporarily ambiguous pronouns and how these were resolved within a critical time window for an anticipatory effect.

8.1.1 Prediction of upcoming reference

Coreference in a discourse is typically expressed through anaphors, very often personal pronouns, to avoid that an expression is repeatedly re-mentioned. In the absence of a cue like gender and/or number marking, pronouns are often, at least temporarily, ambiguous. The interpretation of ambiguous pronouns has been argued to be influenced by several factors like order-of-mention (based on the Structure Building Framework by Gernsbacher & Hargreaves, 1988), subjecthood (e.g., Crawley, Stevenson, & Kleinman, 1990; Frederiksen, 1981) or the preference to keep grammatical roles in parallel (based on the Parallel Function Hypothesis; see, e.g., Sheldon, 1974; Smyth, 1994); for an overview, see Järvikivi, van Gompel, Hyönä, and Bertram (2005). Information structure has been claimed to play a role as well, for example, in the Form Specific Multiple Constraint Approach by Kaiser and Trueswell (2008). Often these factors are confounded and it is difficult to assign an observed preference to only one factor, as shown in example (87) below.

(87) **Sheldon** talks to Lennard. **He** (NP1, subject, topic) is upset that Lennard does not respect his bathroom schedule.

According to Kehler et al. (2008), the crucial influencing factor when establishing coreference is the coherence-relation. They claim that previously observed pronoun interpretation or re-mention biases can be explained in terms of the probabilistic expectation as to which entity will be mentioned next conditioned on the coherence relation that is expected to follow. Thus, they assume that two types of probabilistic information interact: First, the language user expects a certain coherence relation. Second, conditioned on the occurrence of that coherence relation, she expects a pronoun to corefer with an antecedent with a particular grammatical or thematic role. To compute the likelihood that a pronoun will corefer with a certain antecedent in the prior sentence, for example the subject, one has to sum, over all coherence relations, the likelihood of encountering that coherence relation multiplied by the likelihood of a subject reference given that coherence relation. The authors moreover split the bias towards a certain antecedent into two: an expectation regarding the referent (P(referent)) and an expectation about the form of the referring expression the language user would choose to mention that referent (P(pronoun|referent)). This way, they can account for those cases in which, for example, a (3rd person, unaccented) pronoun is used to refer to a non-subject referent, although non-subjects are less often referred to by a pronoun, if there is a suitably large re-mention bias towards the non-subject referent. This approach towards coreference establishment highlights the predictive component that is involved during discourse processing, making use of the Bayes' rule introduced in subsection 2.2.2.

8.1.2 Implicit causality and its effect on coreference establishment

A further factor that can influence pronoun interpretation is the semantic bias of verbs. As an illustration, see the two examples below: Although in both sentences the famous Big Bang Theory character *Sheldon* is the subject of the sentence, only in (88-a) is he the causer or Stimulus of the event or emotion, whereas in (88-b) he is the Experiencer of it. When now adding a subordinate clause with a causal connective, the Experiencer-object verb *frighten* exerts a strong bias to resolve the upcoming pronoun towards the NP1 or subject antecedent, whereas the Experiencer-subject verb *fear* exerts a bias towards the NP2 or object-antecedent.

(88) a. **Sheldon** frightens Lennard because **he** …
b. Sheldon fears **Lennard** because **he** …

In the following, I will describe this bias, known as the implicit causality (IC) bias, and its influence on coreference establishment in more detail, before showing how it can influence sentence interpretation in an anticipatory fashion.

It is well known that IC can affect people's expectations about upcoming (co)reference. The bias is associated with certain verb classes; above it was shown for psych verbs, but also some action verbs (e.g., *to apologize*) have been shown to exhibit an IC bias (e.g., Ferstl, Garnham, & Manouilidou, 2011). Investigation started in the 1970s with Garvey and Caramazza (1974) and Garvey, Caramazza, and Yates (1976), and many researchers since then have investigated the phenomenon of IC, not only in English (e.g., Brown & Fish, 1983; Ferstl et al., 2011), but also in German (e.g., Bott & Solstad, 2014; Rudolph & Försterling, 1997), Dutch (e.g., Cozijn et al., 2011; van Berkum, Koornneef, Otten, & Nieuwland, 2007), French (Guerry, Gimenes, Caplan, & Rigalleau, 2006), Spanish (Goikoetxea, Pascual, & Acha, 2008), Italian (e.g., Mannetti & de Grada, 1991), Norwegian (Bott & Solstad, 2014), Finnish (Järvikivi et al., 2017; Pyykkönen & Järvikivi, 2010), Afrikaans (Vorster, 1984), and Korean (Kim & Grüter, 2019), as well as Japanese, Chinese and Russian (e.g., Hartshorne, Sudo, & Uruwashi, 2013). According to the most recent account developed by Hartshorne and Snedeker (2013) and Hartshorne (2014), the Semantic Structure Account, it is the argument and discourse structure that constitutes the IC bias. As can be seen in their example sentences given in (89) and (90), the argument structure of the psych verbs *frighten* and *fear* favors either the NP1 or the NP2 as the causer of the event (E), as shown in the predicate decomposition schema below, and, thus, as the referent of the ambiguous pronoun.

(89) **Sally** frightened Mary because **she** is scary.
cause (NP1, E) emotional state (result(E), emotion, NP2)

(90) Sally feared **Mar**y because **she** is scary.
emotional state (E, emotion, NP1) in reaction to (E, NP2)

The involvement of discourse structure becomes evident when changing the connective and with it the coherence relation to a Result relation, like in (91). Now, the preferred referent of the pronoun is the Experiencer argument *Mary*. In contexts like these, researchers have used the term *implicit consequentiality* as another form of re-mention bias (e.g., Au, 1986; Crinean & Garnham, 2006; Pickering & Majid, 2007).

(91) Sally frightened **Mary** so **she** ran away.

Kehler et al. (2008) found that verbs with IC bias elicit a stronger-than-usual expectation for an Explanation relation when compared to non-IC verbs in a sentence completion task with and without the causal connective *because*, which explicitly signals an Explanation relation (for similar findings and potential reasons regarding different explanation patterns for psych and action verbs with IC bias due to different verb semantic properties, see Bott & Solstad, 2014). The influencing role of other connectives and, thus, different coherence relations in combination with IC verbs is a controversial issue (see, e.g., Koornneef & Sanders, 2013) and will not be further discussed here.

Hartshorne, O'Donnell, and Tenenbaum (2015) describe the IC bias as rather explicit, with the information about causality and affectedness already encoded in the verb. The authors argue against an assumption that re-mention biases can be explained in terms of inference from general world knowledge, like human thinking about the causality of events (e.g., Brown & Fish, 1983; Rudolph, 2008) or gender stereotypes (for corpora of IC verbs that consider gender, see Ferstl et al., 2011; Goikoetxea et al., 2008).[36] Hartshorne, O'Donnell, and Tenenbaum also point out, however, that further pragmatic inference is needed to refer to the causer or the affected entity, the Experiencer, and that this inference might turn out to be incorrect as in (92).

(92) Sally frightened **Mary** because **she** is so timid.

As support for their claim that the underlying linguistic structure forms the IC bias, Hartshorne, O'Donnell, and Tenenbaum (2015) refer to online studies demonstrating that IC information rapidly affects the interpretation of a pronoun. Whereas an immediate effect of the lexical-semantics of verbs has been shown before, as in the often cited study by Altmann and Kamide (1999), see also Experiment 1, it would be less clear how quickly people are able to infer the cause or consequence of an event on the basis of general world knowledge. Similarly, Koornneef and van Berkum (2006, p. 461) argue that even if IC verbs happen to express interpersonal exchanges, from a language processing perspective there is nothing special about these verbs when compared to a verb like *eat* that restricts the domain of subsequent reference to edible objects.[37]

Using time-sensitive measures, several studies have demonstrated that IC information has an immediate effect on sentence interpretation in adult L1 speakers. Table 8.1 presents an overview of experiments testing the use of IC information in online sentence processing. In probe recognition tasks a probe word, in the current examples the names of the referents, were presented to the participants at specific positions in a sentence and they had to indicate whether the name previously appeared or not; their response latency was then taken as an indication of processing delay or advantage.

[36] One such view on the IC bias is that causal attributions result from perceived covariation patterns, a hypothesis first put foward by Brown and Fish (1983). According to this hypothesis, the IC bias can be explained in terms of perceived consensus and distinctiveness. For example, in case of an NP1-biasing Experiencer-object verb like *frighten* just a few people would be perceived as the causer (low consensus) but many people could be the Experiencer of that event (low distinctiveness).

[37] For me, this does not necessarily rule out that general world knowledge plays a role as well. However, general world knowledge seems to be expendable if the linguistic structure of the verb already provides the information needed. It has to be noted, however, that the argumentation is inconclusive in so far as Altmann and Kamide (1999) did not exclusively test lexical-semantics.

Study	Language	Method	Findings
Greene and McKoon (1995)	English	Probe recognition	faster response times for a test word presented after the verb and connective but before the pronoun compared to a prior probe position when it matches with the bias, effect is confined to NP2-biasing verbs (explained in terms of a countervailing recency effect)
McDonald and MacWhinney (1995)	English	Cross-modal probe recognition	first-mention advantage for NP2-biasing verbs 100 ms after mentioning the 2nd referent, which is canceled out at the pronoun, returns 200 ms later and disappears again at sentence end
Long and de Ley (2000)	English	Probe recognition	slower responses to NP1 name for probes after the pronoun in NP2-bias contexts but only for skilled readers, same effect for congruent endings in Exp.3 for sentence end probe; first-mention advantage for earlier probes and faster response times for probes after congruent endings irrespective of match between name and referent of the pronoun for less skilled readers
Koornneef and van Berkum (2006)	Dutch	Self-paced reading	reading delay after a bias-incongruent pronoun on the following two words
		Eye-tracking during reading	effect in first pass measures at and shortly after a bias-incongruent pronoun
van Berkum et al. (2007)	Dutch	ERP	P600 after a bias-incongruent pronoun
Pyykkönen and Järvikivi (2010)	Finnish	Visual-world eye-tracking	effect of IC after verb onset, even before onset of the causal connective
Featherstone and Sturt (2010)	English	Eye-tracking during reading	effect in first pass measures on the word following a bias-incongruent pronoun
Cozijn et al. (2011)	Dutch	Visual-world eye-tracking	effect of IC before disambiguating information is encountered
Järvikivi et al. (2017)	Finnish	Visual-world eye-tracking	effect of IC before disambiguating information is encountered; first-mention preference does not precede effect of IC; effect of IC for personal and demonstrative pronouns

Table 8.1: Overview of online studies that found an early effect of implicit causality information. The last column briefly summarizes the main findings.

In the cross-modal probe recognition paradigm, sentences are presented aurally and the probe word is presented visually on a screen. Below, examples from the study by Long and de Ley (2000) are given; the numbers indicate the positions where a probe appeared in the different conditions. The first two experiments only included bias-congruent sentences (93), while the third also manipulated bias-congruency (94).

(93) Kate agitated Jane because 1 she (2) smacked her gum continuously (2).

(94) a. Kate agitated Jane at the movies because she 1 smacked her gum continuously 2.

b. Kate agitated Jane at the movies because she 1 was already feeling anxious 2.

One of the authors' main findings was that IC information seemed to influence pronoun resolution in such a way that the non-referent was suppressed in NP2-bias contexts. Other experimental studies employing the method of self-paced reading, eye-tracking during reading or ERP recordings have shown that when participants encountered a pronoun incongruent with the verb's bias they exhibited reading delays or a P600. Visual-world eye-tracking studies have demonstrated an effect of IC that was reflected by anticipatory eye-movements towards the biased antecedent. A more detailed description is given further below. Cozijn et al. (2011), who presented participants with disambiguating information towards the end of their critical sentences (e.g., *x bored y [. . .] because he had told/heard the story already*), additionally found a congruency effect in the first of their two experiments: Participants looked more often and sooner at the referent in the congruent than in the incongruent condition. Guerry et al. (2006), who used a timed story continuation task, further found that it took participants longer to imagine a continuation for a sentence with an anaphor that was incongruent with the verb's bias.

Whereas some researchers have argued in favor of the Clausal Integration Account (Garnham, 2001; Stewart, Pickering, & Sanford, 2000), which states that IC information is used retroactively towards the end of a sentence when both clauses are integrated to form one interpretation, far more studies provide support for the so called Immediate Focusing Account. As already indicated by its name, this account states that IC information immediately exerts its effect on sentence processing. Proponents of this account, however, disagree on the exact underlying mechanisms. Based on findings from experiments using a probe recognition task, McKoon, Green, and Ratcliff (1993) explained the immediate effect of IC in terms of the relative accessibility of the referents in the language user's discourse model, with one discourse referent being more focused as opposed to the other one, leading to facilitation at a bias-congruent pronoun.[38] A more forward-looking mechanism was proposed by Koornneef and van Berkum (2006). According to them, the semantics of the verb, particularly together with a connective like *because*, can be used proactively to predict how the discourse continues. They further suggest that even a specific pronoun might be anticipated. In their eye-tracking during reading experiment, Featherstone and Sturt (2010) tested the latter claim and investigated whether it was a specific word form or reference in general being predicted by also using the bias-neutral word *there* at the beginning of the subordinate clause. Their findings also showed a delay immediately after a bias-incongruent pronoun (95-b) but not after *there* (95-c), providing no support for word-specific prediction. Featherstone and Sturt do not see the mechanism of focusing as an alternative to referential prediction but suggest that both might be involved

[38]In these experiments the probe was only presented after the critical sentence. The results showed faster response times after bias-congruent continuations for a probe (character name) when it matched with the antecedent of the pronoun.

in the processing of IC information.

(95) a. Ryan charmed Emma, because **he** had recently been much more well-behaved, and seemed to have become a gentleman.

b. Emma charmed Ryan, because **he** had recently been much more attracted to women with great personalities.

c. Ryan charmed Emma, because **there** had recently been much more romance in their otherwise unremarkable relationship.

The two accounts mentioned, the Immediate Focusing Account and the Clausal Integration Account, do not necessarily contradict each other. Thus, Koornneef and van Berkum (2006) see their findings as compatible with an Incremental Clausal Integration Account in which "the information made available by the subordinate clause is 'retroactively' related to the interpretation of the main clause on a word-by-word basis" (p. 459); see also Kehler et al. (2008, p. 36f.). As further discussed below, although IC information can be used immediately (even predictively), IC information and other information, for example order-of-mention, have been shown to play a role later on.

8.1.3 The predictive use of implicit causality information

To better answer when and how IC information affects processing, I will now focus on the visual-world eye-tracking studies that used temporarily ambiguous sentences. Several experiments conducted in Finnish and Dutch demonstrate that IC information can be used predictively, meaning before disambiguating information is given, as suggested by Koornneef and van Berkum (2006).

In a study by Pyykkönen and Järvikivi (2010), after a preview of 50 ms participants were presented with two-sentence contexts in Finnish as in (96), while they were looking at four pictures like, for the sentence context cited here, a picture of a guitarist, a butler, a dining room and a stage. Their task after some trials was to continue the stories they had just heard. The authors' interest was to detect when IC information was activated, before or after the pronoun, and how IC information interacted with other factors, especially structural information like order-of-mention, grammatical role and parallel syntactic function of the pronoun and the potential antecedent. Therefore, they also manipulated the pronoun type, i.e., whether it was a subject, as in (96-a), or an object, as in (96-b). In Finnish, it is grammatical to start a clause with an object pronoun. Moreover, due to its properties – all verbs are morphologically complex – all of the verbs that appeared in the same context shared the same root.

(96) Kitaristi oli valmistautumassa illan esitykseen. Hovimestari **pelotti/pelkäsi**
the guitarist was preparing for the night performances the butler **frightened/feared**
kitaristia ravintolasalissa, koska koko päivän . . .
the guitarist in the dining room, because all day . . .

'The guitarist was preparing for the night's performance. The butler frightened/feared the guitarist in the dining room because for the whole day ... '

a. **hän** kummallista kyllä oli näyttänyt erittäin tyytymättömältä huolimata tulossa
he curiously enough was seemed to very grumpy despite pending
olevasta suositusta illasta.
the special evening
'he curiously enough had seemed extremely unhappy despite the upcoming popular night.'

b. **hänet** kummallista kyllä oli näyttänyt erittäin tyytymättömälä huolimata tulossa
him curiously enough was seemed to very grumpy despite pending
olevasta suositusta illasta.
the special evening
'he curiously enough had been seen (to be) extremely unhappy despite the upcoming popular night.'

For the time windows before pronoun onset, the authors found an effect of IC information that emerged even before the connective was encountered: Participants showed more fixations on the subject for NP1-biasing verbs and more fixations on the object for NP2-biasing verbs consistent with the IC bias. The main effect of IC was also significant in the 300 ms time window after connective onset. For the time window after pronoun onset, a main effect of IC consistency emerged in the first 300 ms time window during the pronoun and reappeared in the fourth time window from 900–1200 ms, which also showed an interaction between pronoun type and grammatical role, indicating that there were more looks to the object antecedent for the object pronoun *hänet* than for the subject pronoun *hän* and vice versa. In the 600–900 ms time window, there was a marginally significant interaction between IC consistency and grammatical role, indicating that the effect was temporarily more pronounced for the subject antecedent. In all time windows the researchers found a main effect of grammatical role, with more looks to the subject antecedent relative to the object antecedent. The conclusion Pyykkönen and Järvikivi drew from these findings was that IC information can be activated even before the explicit trigger, the causal conjunction, but it has no privileged role as a pronoun resolution cue throughout.

In a study by Cozijn et al. (2011), the participants listened to sentence contexts in Dutch, while they saw three line drawings, the two referents at the bottom to the left and right side of the screen facing each other, and a distractor, for the example (97) below a car, which was displayed above in the center. In a first experiment, participants had to name the referent of the pronoun at the end of each trial; in a second experiment they had to verify statements testing their comprehension. Moreover, the second experiment used filler sentences without IC verbs and connectives other than *omdat* (because), and the neutral clause in the sentence contexts that was intended to extend the critical region in the first experiment was removed. There was no preview time. In (97-a), a bias-congruent sentence continuation is given, and in (97-b) a bias-incongruent one. The digits indicate (natural) pauses. In the second experiment with a higher speaking rate, a standardized clausal pause of 600 ms was inserted between the main and subordinate clauses.

(97) De octopus verveelde de krokodil in de auto 1 [omdat hij rusteloss was$_{\text{neutral clause}}$ 2]
The octopus bored the crocodile in the car 1 [because he restless was$_{\text{neutral clause}}$ 2]

a. en omdat hij het verhaal al zeker tien keer had **verteld** tijdens de rit.
and because he the story already at least ten times had told during the ride.
'and because he had told the story already at least ten times during the ride.'

b. en omdat hij het verhaal al zeker tien keer had **gehoord** tijdens de rit.
and because he the story already at least ten times had heard during the ride
'and because he had heard the story already at least ten times during the ride.'

For the analysis, the sentences were divided into single segments, whereby the on- and offset of each segment was shifted 200 ms forwards. In the first experiment, IC information became effective in the first connective and pronoun segment and the rest of the neutral clause for NP1-biasing verbs (interaction between bias and NP-area). In the first pause segment, the participants showed a preference for the NP1 referent. According to the authors, this NP1-preference might have canceled out an effect for NP2-biasing verbs. In the second experiment, the effect of IC emerged in the segment after the connective and pronoun but before the disambiguating word for both verb types. No NP1-preference in the pause segment was found here. Crucially, this study demonstrated that Dutch L1 speakers were more likely to fixate the biased antecedent before disambiguating information was encountered.

Another study by Järvikivi et al. (2017), including two experiments, tested sentences like (98). The first experiment aimed at unraveling the effect of grammatical role and order-of-mention, while the second compared the effect of IC information for different referential expressions, the personal pronoun *hän* and the demonstrative pronoun *tämä* in Finnish. In both experiments the preview time was 50 ms. As in Pyykkönen and Järvikivi (2010), the task was to continue some of the stories.

(98) a. Vladimir Putin peloti/pelkäsi George Bushia Valkoissessa talossa
Vladimir Putin-*nom-sub* frightened/feared George Bush-*ptv-obj* at the White House
koska hän ...
because he ...

b. Vladimir Putinia peloti/pelkäsi George Bush Valkoissessa talossa
Vladimir Putin-*ptv-obj* frightened/feared George Bush-*nom-subj* at the White House
koska hän ...
because he ...

In the first experiment, the sentences appeared in the four experimental conditions as shown in the examples above, i.e., SVO, NP1-bias or SVO, NP2-bias (98-a) and OVS, NP1-bias or OVS, NP2-bias (98-b). In addition, all sentences also appeared in a counterbalancing condition where the proper names (e.g., Vladimir Putin and George Bush) were swapped. For the OVS sentences, the object in partitive case was topicalized. A pause of 600 ms was inserted between the main and subordinate clauses. Participants would see four pictures, the two referents, a mentioned location and a distractor picture showing another character that was not mentioned. The final statistical analyses took into account fixations from 300 ms after pronoun onset. The results of their analyses showed an effect of IC congruency/consistency that was marginal in the time window 900–1100 ms and became significant for the two later time windows plus an effect of order-of-mention in the time windows 1100–1300 ms (significant by items but marginal by participants) and 1300–1500 ms. There was no effect of grammatical role. Up to 1500 ms the sentences were semantically neutral regarding the interpretation of the pronoun. Järvikivi et al. take the finding that the effect of IC preceded that of order-of-mention as evidence for immediate focusing. Moreover, they assume that the previously observed subject preference resulted from a semantic component, insofar as the subject role is often associated with agentivity (see also Schumacher, Roberts, & Järvikivi, 2017).

The second experiment contrasted the 3rd person personal pronoun *hän* with the demonstrative pronoun *tämä* (this). This time only SVO structures were used; see (98-a) above. Thus, the conditions became NP1-bias, *hän* vs. NP2-bias, *hän* vs. NP1-bias, *tämä* vs. NP2-bias, *tämä*. Again, all sentences appeared in a counterbalanced version with the character positions reversed. The results showed a main effect of IC bias in the time windows 300–500 ms (only by participants), 500–700 ms and 700–900 ms, as well as 900–1100 ms (only by items) after pronoun onset. An interaction between order-of-mention and type of pronoun starting 500 ms (significant by items and participants from 700 ms onwards) after its onset indicated a first-mention preference for *hän* and a second-mention preference for *tämä* (confirmed through separate analyses). The effect of IC was thus visible for both referential expressions, the pronoun and the demonstrative, however both seemed to be differently affected by structural prominence.

To summarize, all studies described above found an effect of IC information that emerged before disambiguating information was encountered, thus demonstrating prediction according to the criterion formulated in the beginning of this thesis. The studies moreover showed an effect of grammatical role, which influenced the results insofar as the subject, in the experiments by Pyykkönen and Järvikivi (2010) and Cozijn et al. (2011) also the NP1, was favored over the object antecedent. However, the results of the first experiment by Järvikivi et al. (2017) indicate that this preference might be the result of the subject's semantic properties, because subjects are typically associated with agentivity. Their study further showed that order-of-mention played a role but only after IC information exerted an effect.

8.1.4 The use of implicit causality information by different participant groups

There is indication that even within L1 speakers the use of IC information can vary. Long and de Ley (2000), who in their probe recognition task tested different types of readers, found that less skilled readers demonstrated a first-mention/subject advantage: Irrespective of the IC bias of the verb, responses to NP1 names were faster than to NP2 names. Less skilled readers, moreover, showed a congruency effect in terms of faster responses after bias-congruent continuations at the end of a sentence. Thus, their response pattern was different from that of the skilled readers, who demonstrated a matching effect, i.e., slower responses to NP1 names in NP2-bias contexts (see overview in Table 8.1). Interestingly, less skilled readers were also university students but were categorized as less skilled relative to their peers based on their performance in a vocabulary and comprehension section of a reading test.

Cheng and Almor (2017), as described in subsection 4.2.6, used a sentence completion task to examine the use of implicit causality and implicit consequentiality information in L1 and L2 speakers of English. Similar to the less skilled readers in Long and de Ley, the intermediate-advanced and advanced L2 speakers demonstrated an overall NP1-preference. This preference, which was also seen in the implicit consequentiality experiment, might explain why the effect was only significantly reduced for NP2-biasing verbs in the IC information experiment. Cheng and Almor (2017) suspected that the reason could be a generally reduced ability to generate expectations in L2 speakers as claimed by the RAGE hypothesis. More recently, Cheng and Almor (2019) argued that the difference observed between these L1 and L2 groups results from different beliefs about pronoun use. When asking the groups to produce a sentence continuation without a pronoun prompt, no distinct pattern for L2 speakers was observed; a distinct pattern was only observed when a pronoun was provided. Contemori and Dussias (2018), who tested L2 speakers of English with an earlier AoA, only reported a difference in timing between the L1 and L2 groups. Crucially, both groups used IC information predictively, although only for NP1-biasing verbs.

Previous studies with children are difficult to interpret insofar as they did not use the same experimental design as for the adult group they were compared to like in Au (1986), or did not exclusively test the use of IC information as in Corrigan (1994) or Rudolph (2008). While the adults in Au (1986) conducted a sentence completion task, the children with an age of around five years had to respond to questions as shown in (99). Nevertheless, the results from this offline task showed that English-speaking children were sensitive to the IC bias of the verbs (10 action and 10 psych verbs) they were presented with.

(99) This is John and this is Mary. John thanked Mary. Can you guess why? Why did John thank Mary?

Findings from an experiment conducted by Corrigan (1994) with three- to five-year-old English-speaking children demonstrated causal events with the help of toys. The children were required to produce a response including the verb involved, for example *to push*. The authors also manipulated animacy patterns and the status of the participants (e.g., mother-child) as they were more interested in a link to social-cognition and language. Experiments on German conducted by Rudolph (2008) rather tested for perceived covariation patterns, i.e., the perception of consensus and distinctiveness.

Results from a sentence completion task by Goikoetxea et al. (2008) with the aim of providing normative data on IC verbs in Spanish showed a similar bias for children at the age of eight to nine years and 11 to 13 years. However, the NP1-bias was stronger in the adult group and, in addition to the interaction between verb bias and age, there was an interaction between verb type and age. For the state verbs (here referred to as psych verbs), the children tended to refer to the NP2 more often than the adults. The last finding was explained by the authors in terms of a recency effect. Another noteworthy study, although not on the IC bias but the acquisition of emotion verbs, was conducted by Hartshorne, Pogue, and Snedeker (2015). The authors tested English children's understanding of verbs like *frighten* and *fear* in several experiments using a truth-value judgment task. Although more frequent as calculated from frequencies in child directed speech from the CHILDES database (MacWhinney, 2000), *fear*-type verbs such as *like* or *love* were more difficult to understand than verbs that follow the canonical order with the first-mentioned subject being the causer. By four years of age, children were beginning to understand sentences involving *frighten*-type verbs. The authors see these findings in line with the assumption that children recognize and exploit the mapping between caused events and transitive syntax in acquisition. Why children are nevertheless found to produce sentences involving *fear*-type verbs is explained by the authors in terms of potential alternative strategies like placing the animate argument first. So far, there is no study I know of that has tested how children process IC information online.

8.1.5 The current study

Experiment 4 investigated how adult Russian L1-German L2 speakers and German-speaking children process IC information. The results of the learner groups were not only compared to each other but also to a group of adult L1 speakers as a reference group. The central question was whether IC information was used predictively by the learner groups as seen in adult L1 speakers in previous studies (e.g., Cozijn et al., 2011; Järvikivi et al., 2017; Pyykkönen & Järvikivi, 2010) and, if there were differences, whether these were qualitative in nature. The results of previous offline sentence completion tasks have shown a similar bias in children, although they exhibited a tendency to refer to the most recent antecedent for psych verbs (Goikoetxea et al., 2008), but a reduced bias and overall NP1-preference for late L2 learners (Cheng & Almor, 2017). Cheng and Almor (2017) hypothesized that a potential reason could be L2 speakers' reduced ability to predict. Experiment 4 aimed to find out if there was indeed a difference between L1 and L2 processing in terms of reduced prediction at the discourse-level in L2 speakers (Cheng & Almor, 2017; Grüter et al., 2014, 2017). If so, we expected a difference in the processing time course of adult and child L1 speakers and adult L2 speakers.

On the other hand, a learner pattern could be expected. An eye-tracking study by Pozzan and Trueswell (2016) that required participants to act out a task (e.g., *Put the frog on the napkin onto the* *box*) showed that for a phenomenon that potentially requires more cognitive resources, in their study sentence revision, child L1 and adult L2 speakers showed a similar act-out pattern, although the L2 speakers' prior looking pattern resembled that of adult L1 speakers. The authors attributed this difficulty with revision to an enhanced cognitive load in the L2 group. Several researchers have proposed that prediction also hinges on the availability of cognitive resources (e.g., Huettig, 2015; Kuperberg & Jaeger, 2016). Huettig and Janse (2016) found that individuals' working memory and speed of processing mediated anticipatory eye-movements in visual-world eye-tracking (see subsection 3.2.2). In the case of discourse-level prediction based on IC information, which requires the integration of argument and discourse structure, we might expect that children and L2 speakers pattern alike. If, however, prediction, unlike sentence revision, in children is not or is less affected by cognitive resources, a similar pattern in child and adult L1 speakers would be expected. Testing children, who are cognitively less mature, and adults who have to process a later learned language, which presumably is cognitively more demanding, makes it possible to account for an effect of cognitive resources. Here this was done by taking into account memory capacity, as this might be a crucial mediating factor for prediction and, in the current experiment, information had to be retrieved. Thus, the comparison between these two different learner groups tested for another potential origin of reduced prediction in L2 speakers, i.e., a cognitive origin.

The crucial research question was whether participants' probability to fixate the biased antecedent significantly increased before the disambiguating information was encountered. A further interest was to determine the onset of the effect, i.e., when there was a significant difference between NP1- and NP2-biasing verbs for the probability that the NP1 referent was fixated. The current study tried to answer the following research questions:

RQ 1: Do child L1 and adult L2 speakers use IC information in the same way as adult L1 speakers to predict upcoming reference?

RQ 2: Do the three groups differ in the onset of the (anticipatory) effect?

Based on prior research, the adult L1 speakers of German were expected to show an anticipatory effect in terms of a higher probability of fixations on the biased antecedent compared to the non-biased antecedent before disambiguating information was encountered. For L1 children (basic) predictive abilities have been shown at a young age, however prediction based on IC information requires the integration of different information sources and might be more demanding for children than anticipating an upcoming noun based on information about the Agent and the lexical-semantics of the verb (Borovsky et al., 2012) or case marking and verb semantics (Özge et al., 2016). Thus, for the child L1 and adult L2 group expectations were less clear. One might expect both groups to have less experience with German IC verbs than adult L1 speakers. However, as discussed in section 3.1, their type of exposure might also be different. Another difference between these groups was their cognitive maturity. Below, potential outcomes and their interpretation are discussed:

1. If the results show an anticipatory effect for the L1 groups, but not for the adult L2 group, we may take this as evidence that the predictive processing abilities of L2 speakers are indeed reduced as proposed by the RAGE hypothesis, at least for prediction at the discourse-level based on IC information. Such an outcome would further provide an indication that prediction in children might be less dependent on cognitive resources, unlike sentence revision.

2. If the results show an anticipatory effect for the adult L1 group, but not for the learner groups, we may interpret this as a learner pattern and as an indication that prediction based on IC information is cognitively demanding. This finding would also be in line with the RAGE hypothesis.

3. Another possibility is that the adult groups but not the children show an effect of IC information. This would strengthen the assumption that prediction requires the availability of cognitive resources and might be dependent on the complexity of the predictive cue.

8.2 Pre-test 1: Selection of implicit causality verbs

The visual-world eye-tracking experiment was preceded by two pre-tests. The purpose of the first pre-test described in this section was to find a suitable set of German verbs with IC bias.

8.2.1 Materials & methods

8.2.1.1 Participants

Thirty-five German native speakers (mean age: 22.49, age range: 18–48, 30 female) participated in this pretest. One additional participant who reported being bilingual was excluded. Participants were either volunteers or students at the University of Potsdam, who received course credit for their participation.

8.2.1.2 Design & materials

A total of 18 NP1-biasing Experiencer-object and 18 NP2-biasing Experiencer-subject German psych verbs were selected. These were taken from Bott and Solstad (2014) and Härtl (2001) or were German translations of psych verbs which had a strong IC bias in English according to Ferstl et al. (2011). Apart from the IC bias, their frequency was considered. For this, frequency counts of the German Wortschatz Project (http://wortschatz.uni-leipzig.de/) were used. For verbs which had a similar meaning, the one with the highest frequency was chosen. Furthermore, it was considered whether verbs were appropriate for an experiment with children. This led to the prior exclusion of verbs like *faszinieren* (to fascinate) or *provozieren* (to provoke).

The experimental items were of the form *NP1 verb NP2 connective pronoun [. . .]*. The NPs were always male characters, so the pronoun was ambiguous in German. Common proper names were used that were clearly male and always began with different letters and were of the same length. To avoid interference from semantics or pragmatics as far as possible, legal pseudowords were created to describe the characters in the subsequent subordinate clause. Additionally, every pseudoword in the experimental materials appeared twice, once with an NP1- and once with an NP2-biasing verb. An example item is given in (100).

(100) Lars fürchtet Nico, weil er ein Binsstieler ist. (Wer ist ein Binsstieler?)
Lars fears Nico because he is a PSEUDOWORD. (Who is a PSEUDOWORD?)

Moreover, 14 filler sentences with a different sentence structure than the experimental items, also including female characters, were constructed. All filler sentences contained a verb of transfer and a pseudoword. Participants were asked which of the characters is in possession of this novel object (e.g., *Justus sends Karolin a* PSEUDOWORD. *Who has a* PSEUDOWORD*?*).

8.2.1.3 Procedure

The experiment was set up as a web-based forced-choice questionnaire and took around 15 to 20 minutes. Participants were informed that the pseudowords in the sentences were not important for the underlying research question, they only had to make a decision concerning the characters. Whether a character appeared in the first or second position in the response fields, see Figure 8.1 for an example, was randomly assigned by the experimental software (SoSci survey; Leiner, 2014).

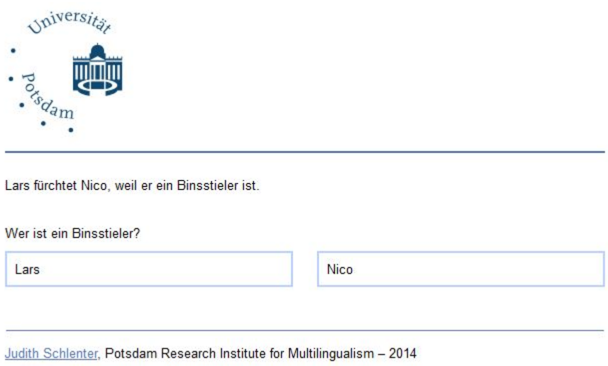


Figure 8.1: Example screen in Pre-test 1. Participant had to choose between two referents.

The 50 sentences appeared in a previously pseudo-randomized order with no more than two verbs of the same type and no more than four experimental items in succession. The same pseudowords never appeared consecutively. Only one item was presented at a time. It was not possible to return to the previous screen, of which the participants were informed beforehand. Two presentation lists were constructed by reversing the order of the first one.

8.2.2 Results

Table 8.2 shows how often the expected referent was chosen, indicating the bias strength of a verb. Verbs that were used in the material construction for Experiment 4 are in bold.

Experiencer-object verb		NP1 decisions in %	**Experiencer-subject verb**		NP2 decisions in %
empören	**outrage**	94.29	**bedauern**	**be sorry for**	94.29
amüsieren	amuse	91.43	**vergöttern**	**adore**	94.29
beeindrucken	**impress**	91.43	**bemitleiden**	**pity**	91.43
begeistern	**inspire**	91.43	**mögen**	**like**	91.43
beunruhigen	**agitate**	91.43	**fürchten**	**fear**	88.57
erschrecken	**scare**	91.43	**bewundern**	**admire**	85.71
verwirren	**confuse**	91.43	**hassen**	**hate**	85.71
schockieren	shock	88.57	schätzen	appreciate	85.71
überraschen	**surprise**	88.57	**verachten**	**despise**	85.71
ängstigen	**frighten**	85.71	vertrauen	trust	85.71
erfreuen	**delight**	85.71	**bevorzugen**	**prefer**	82.86
gruseln	give the creeps	85.71	**lieben**	**love**	82.86
verärgern	**anger**	85.71	misstrauen	distrust	82.86
verblüffen	**amaze**	85.71	**respektieren**	**respect**	80.00
langweilen	bore	82.86	verdächtigen	suspect	80.00
erstaunen	**astonish**	80.00	**verehren**	**worship**	80.00
enttäuschen	disappoint	80.00	beneiden	envy	77.14
beleidigen	insult	22.86	übersehen	overlook	60.00

Table 8.2: Pre-test 1: Bias strength for the German psych verbs included in Experiment 4. Verbs that were used in the materials are in bold.

8.3 Pre-test 2: Material evaluation

The second pre-test tested the materials designed for the visual-world eye-tracking experiment and, in particular, assessed whether the materials were appropriate for an experiment with child L1 and adult L2 comprehenders. At the same time, the pre-test assessed participants' offline sensitivity to IC information.

8.3.1 Materials & methods

8.3.1.1 Participants

The adult L1 group consisted of 24 German native speakers (15 female, 2 left-handed) with a mean age of 24.92 years (range: 18–34), none of whom reported being early bilingual. The child L1 group consisted of 35 German-speaking children (13 female, 2 left-handed). Their mean age was 11 years and two months (range: 8;6–13;0). The age was chosen based on the expectation that children at this age already knew all the critical verbs and were able to pay attention during the course of the experiment (similar to the age range tested by Goikoetxea et al., 2008). Two additional children were subsequently excluded, one because she was much younger than the other children and another because she produced more than ten timeouts. The adult L2 group consisted of 31 L2 speakers of German with Russian as L1 (27 female, 4 left-handed) with a mean age of 27.42 years (range: 20–36). All L2 learners included were highly proficient in German as measured by the Goethe placement test (Goethe-Institut, 2011), assigning them to C1- to C2-level according to the CEFR (Council of Europe, 2001). The mean Goethe score was 26.55 (SD: 2.12, range: 22–30). One additional participant was tested but excluded due to lower proficiency. All of the L2 speakers were currently residing in Germany and their average time of exposure to German was 13.94 years (SD: 5.16, range: 5–24).

All adult participants received either course credit or a payment of four euros. The children received a payment of four euros and a sticker. All participants gave informed written consent regarding the data use, in the case of the children this was done by a caretaker. Ethical approval for experiments with children for the Potsdam Research Institute for Multilingualism was received from the ethics committee of the University of Potsdam (37/2011).

8.3.1.2 Design & materials

For the experimental items, 48 sentences like in (101) were constructed. These included 24 psych verbs from the first pre-test, half of them biasing towards the NP1 and the other half biasing towards the NP2. All verbs were used twice. Twelve different combinations of depictable male characters were used, which appeared in the reversed order the second time they were mentioned with the same verb type in another context. The experimental items included no female characters, because the pronoun *sie* (she) is ambiguous between third person singular feminine and third person plural in German. Two pictures that showed the potential antecedents of the pronoun were presented to the participants (Figure 8.2). To be able to measure the bias strength for the experimental sentence contexts, the sentences and scenes were globally ambiguous.

(101) Der Indianer ängstigt den Cowboy am Lagerfeuer, weil **er** zufällig ein
The Native American frightens the cowboy at the campfire because **he** accidentally a
Messer hervorholt.
knife takes out
'The Native American frightens the cowboy at the campfire because he accidentally takes out a knife.'

Figure 8.2: Visual display for an experimental sentence in the picture selection task of Pre-test 2

In addition, 48 filler sentences were constructed that also included alternative connectives to *weil* (because), contained a different ordering of sentence segments and used female characters and/or plurals; for examples, see (102) and (103). Different from the experimental sentences, there was always a morphosyntactic cue indicating who the referent of the pronoun was.[39] Therefore, the contrast was between the picture showing the correct referent and the picture showing the other referent, who was mentioned but to whom the pronoun did not refer. The fillers were intended to indicate whether participants would correctly resolve a pronoun when an unambiguous cue was given.

(102) **Die Verkäuferin** grüßt freundlich den Bauarbeiter am Stehtisch,
The saleslady-[*fem.*] greets cordially the construction worker-[*masc.*] at the bar table,
als **sie** den Besen abstellt.
while **she** the broom puts aside
'The saleslady cordially greets the construction worker at the bar table, while she puts aside the broom.'

pictures: saleslady with broomstick vs. construction worker with broomstick

(103) **Die Einbrecher** sehen am Zaun die Polizistin, als **sie** durch ein
The burglars see at the fence the policewoman, while **they** through a
Fernglas blicken.
pair of binoculars look-[*pl.*]
'The burglars see the policewoman at the fence while looking through binoculars.'

pictures: 2 burglars with binoculars vs. policewoman with binoculars

All sentences were recorded in a sound attenuated room and spoken by a female native speaker of German at a normal speaking rate. Because the experimental items were supposed to resemble those used

[39] Arnold, Brown-Schmidt, and Trueswell (2007) demonstrated in a VW experiment that a morphosyntactic cue like gender marked on a pronoun can already be used successfully by English-speaking children at the age of three to five years.

in the online experiment, where it was crucial to account for differences in prosody and pace between conditions, the spliced versions for the VW experiment were also used here (for information about the splicing procedure, see subsection 8.4.2). All characters in the sentences were portrayed by photographs of Playmobil® figures. All photos were edited with a picture editing program, so the characters appeared on a white background with a size of 400 x 400 pixels and 72 dpi. The picture selection task was programmed with DMDX (Forster & Forster, 2003). For this, each display, two pictures next to each other, was saved in 24-bit bitmap format. All 96 sentences and the respective pictures were presented to all participants. Items and fillers were mixed in a pseudo-randomized order and presented in two different orderings to counterbalance potential training or fatigue effects. Whether the picture with the biased antecedent appeared on the left or the right side of the screen was also counterbalanced.

8.3.1.3 Procedure

Participants sat in front of a 15.6 inch laptop screen with a resolution of 1366 x 786 pixels. First, they saw a presentation showing all Playmobil® characters that appeared in the picture selection task with their corresponding names. After this familiarization phase, the actual experiment started. While the participants were listening to a sentence, two pictures were presented on the screen. Their task was to indicate which picture best described the event by pressing either the left button for the left picture or the right button for the right picture on a gamepad. There was an equal number of left and right expected responses. Participants were instructed to press the buttons quickly after a beep sound was heard. The beep sound was intended to ensure that all participants listened to the complete sentence. If no response was given, the next trial started automatically after ten seconds. Participants were informed that the characters and events in the sentences would reappear several times in different contexts. There was a short practice block consisting of four trials. Two breaks in between allowed participants to rest. One experimental session lasted approximately 30 minutes.

The children were tested either at the Potsdam Research Institute for Multilingualism or at their homes. During the familiarization phase and the practice block, the experimenter sat next to the child. The children were told that they were going to see some characters that they would hear short stories about in the following experiment. Since it turned out that the children needed more instructions than the adults, the children were given pictures of two of the practice items and were asked why they thought that a particular picture described the event better than the other one. If a child used an incorrect strategy, the experimenter tried to explain again without telling her to listen to the pronoun or the verb. Instead they were told that they need to pay attention to who is doing what in the sentence and the examples were used as a demonstration (e.g., *Look, who is lighting a lantern? That is why the picture in which the pirates have the lantern is better.*). To increase their motivation, positive feedback was displayed on the screen when there was a break (*Well done! This was the 1st/2nd part. Press button 3 to continue!*).[40] The procedure for the adult L2 speakers was the same as for the adult L1 group with the exception that the L2 speakers additionally had to fill in a short vocabulary test after the picture selection task that was intended to check whether they were familiar with the IC verbs.

8.3.2 Results

The adults showed a high accuracy for the fillers: The adult L1 speakers had an overall accuracy of 99%, ranging from 94% to 100%, and the adult L2 speakers an accuracy of 97%, ranging from 79% to 100%. Some of the children had severe problems with the task and produced incorrect responses to more than 25%, which was set as the threshold. The child L1 group was therefore split into comprehenders (n = 24), who showed an adultlike pattern for the unambiguous fillers (mean accuracy: 95%, range: 81%–100%), and non-comprehenders (n = 11). Some children in the non-comprehender group seemed to strategically choose the subject of the matrix sentence (n = 5). It is unclear whether the non-adultlike behavior of the other children resulted from simply not being attentive or from following another strategy. Those children (n = 4) mainly belonged to the first children being tested who received the instructions without picture clarification. Thus, in the following I will concentrate on the results of the subgroup who demonstrated successful coreference establishment.

Whereas the adult L1 group showed the expected bias with only one exception, sometimes a sentence context elicited a less clear bias towards the NP1 or NP2 in the child L1 and adult L2 group, a point I will return to in the discussion. Since participants also reported that the characters included in that sentence could not be easily distinguished and the L2 learners had problems with some of the words included, this sentence together with the three other sentences using this character combination were excluded from the analysis for all groups. No further data had to be removed in the adult L1 group. In the child data, 13 trials were missing due to interruptions or because they were timeouts (1% of the child L1 data). All trials that included an unknown verb, as indicated by the L2 participants in the vocabulary test, were removed on a by-participant basis from the analysis of the experimental items; this included a maximum of four different verbs for one participant. These together with three additional timeouts accounted for 4% removal in the L2 data. The adult L1 group showed an average bias strength of 94% for NP1-biasing verbs, ranging from 82% to 100%, and 86% for NP2-biasing verbs, ranging from 68% to 100%. The child L1 group showed an average bias strength of 83% for NP1-biasing verbs, ranging from 45% to 100%, and 81% for NP2-biasing verbs, ranging from 50% to 100%. The adult L2 group showed an average bias strength of 81% for NP1-biasing verbs, ranging from 36% to 95%, and 77% for NP2-biasing verbs, ranging from 45% to 100%.

[40] The positive feedback was included after eight children had already been tested. The extended instructions were used after ten children had been tested.

To compare the groups, the likelihood of selecting the NP1 referent dependent on verb bias and group was calculated; see Figure 8.3, which shows the proportion of NP1 responses per verb bias condition. A generalized mixed model with treatment contrast coding for verb bias and group was computed. Only the model with random intercepts for subject and item converged. Table 3 shows the output of the model with adult L1 group and NP1-bias as well as adult L1 and NP2-bias as the reference level. The effect of verb bias demonstrates that there was a significant difference in NP1 responses for NP1- and NP2-biasing verbs in the adult L1 group. The effects of group indicate that there were fewer NP1 responses for NP1-biasing verbs in both the child L1 and adult L2 group and more NP1 responses for NP2-biasing verbs than in the adult L1 group. Hence, the difference in NP1 responses between verb bias conditions compared to the adult L1 group was reduced as further indicated by the interaction between group and verb bias.

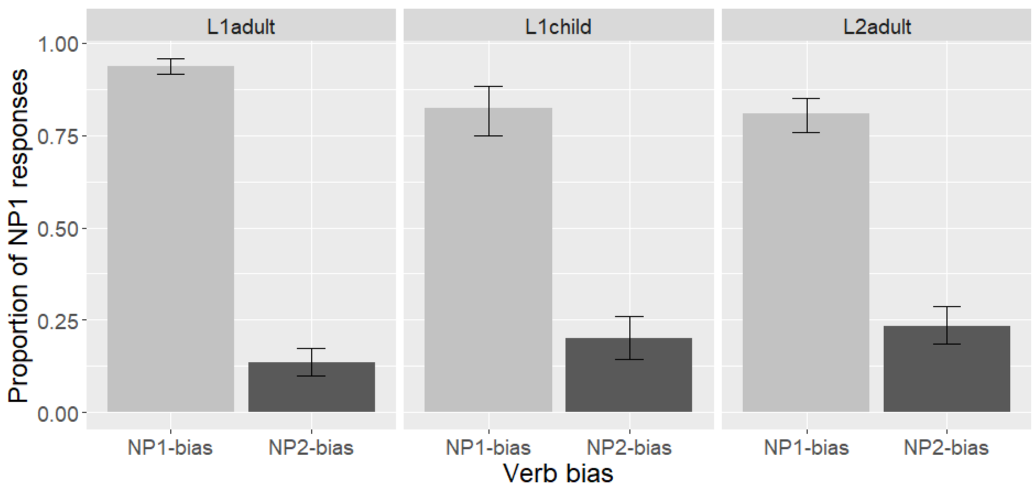


Figure 8.3: Proportion of NP1 responses for NP1- and NP2-biasing verbs for each group in Pre-test 2: adult L1 (left), child L1 (middle) and adult L2 (right). Error bars show the bootstrapped confidence intervals.

To compare the two learner groups, the factor group was re-leveled, so the child L1 group became the reference group. Since already shown, results for the adult L1 group were skipped in Table 8.4. Although no significant difference in the likelihood of NP1 reference for either condition (no effect of group) emerged, the difference in NP1 responses between verb bias conditions turned out to be reduced in the adult L2 compared to the child L1 group as signaled by the interaction between group and verb bias. Finally, the omnibus model was used, and the factor group re-leveled again to test for an effect of verb bias in the adult L2 group. As shown in Table 8.5, the adult L2 group also demonstrated a significant difference in NP1 responses between verb bias conditions. Hence, the results demonstrate an effect of IC information for all groups tested. In Appendix C, additional tables show the total number of NP1 responses for each sentence context per group (Table C.4 and Table C.5) and detailed information about the children (Table C.6).

Reference group: Adult L1	Estimate	Std. Error	*z*-value	*p*-value
Intercept (NP1-bias)	2.8855	0.2216	13.022	<0.001***
Group (L1 child)	-1.1422	0.2369	-4.821	<0.001***
Group (L2 adult)	-1.3283	0.2262	-5.871	<0.001***
Verb bias (NP2-bias)	-4.8253	0.2704	-17.848	<0.001***
Group (L1 child) x Verb bias	1.5443	0.2773	5.568	<0.001***
Group (L2 adult) x Verb bias	1.9958	0.2627	7.598	<0.001***
Intercept (NP2-bias)	-1.9398	0.1781	-10.890	<0.001***
Group (L1 child)	0.4022	0.1951	2.06	0.0393*
Group (L2 adult)	0.6676	0.1819	3.670	<0.001***
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				
Formula: NP1 ~ group*verb bias + (1 | subject) + (1 | item no), family = binomial				

Table 8.3: Pre-test 2: Results of the generalized mixed model (adult L1 as reference group)

Reference group: Child L1	Estimate	Std. Error	*z*-value	*p*-value
Intercept (NP1-bias)	1.7433	0.1739	10.072	<0.001***
Group (L2 adult)	-0.1861	0.1811	-1.028	0.3042
Verb bias (NP2-bias)	-3.2809	0.2245	-14.612	<0.001***
Group (L2 adult) x Verb bias (NP2-bias)	0.4514	0.2168	2.082	0.0374*
Intercept (NP2-bias)	-1.5376	0.1680	-9.154	<0.001***
Group (L2 adult)	0.2653	0.1721	1.542	0.1232
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				
Formula: NP1 ~ group*cond + (1 | subject) + (1 | item no), family = binomial				

Table 8.4: Pre-test 2: Results of the generalized mixed model (child L1 as reference group)

Reference group: Adult L2	Estimate	Std. Error	*z*-value	*p*-value
Intercept (NP1-bias)	1.5572	0.1579	9.861	<0.001***
Verb bias (NP2-bias)	-2.8294	0.2039	-13.876	<0.001***
Intercept (NP2-bias)	-1.2721	0.1517	-8.386	<0.001***
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				
Formula: NP1 ~ group*cond + (1 | subject) + (1 | item no), family = binomial				

Table 8.5: Pre-test 2: Results of the generalized mixed model (adult L2 as reference group)

8.3.3 Discussion

The purpose of this second pre-test was to evaluate the materials designed for a visual-world eye-tracking experiment and to assess participants' offline sensitivity to IC information for these materials. The most important finding was that, in general, all groups were able to use IC information to establish coreference. Nevertheless, there were gradual differences, with the adult L1 group demonstrating a stronger bias than the children in the comprehender group and a stronger bias than the adult L2 group. Children and L2 speakers showed a reduced bias for both verb types compared to the adult L1 group, i.e., fewer NP1 responses for NP1-biasing verbs and more NP1 responses for NP2-biasing verbs. The difficulties observed with pronoun resolution for a subset of the children might be attributed to the complexity of the task. Note that this task is typically used for assessing adults' or children's preferences when there is no correct or incorrect response, like for example in Sekerina et al. (2004). However, overall the adult and most of the child data demonstrated that this task was sufficient to test successful pronoun resolution as shown by the results for the unambiguous fillers.[41]

There was only a single sentence context that induced no clear bias in the adult L1 group. When comparing the preferences for different sentence contexts in the child L1 and adult L2 group, however, some additional sentences turned out to be potentially problematic. One reason could be the order of characters included. It might be that, for example, the burglar was more likely to carry a gun or be the cause of fear than the policeman for some participants. Based on this observation, the order of characters was fully counterbalanced for the visual-world eye-tracking experiment. Most of the vocabulary was known to the adult L2 speakers. Crucially, none of the verbs caused a particular problem. Some nouns that were reported as unknown by several participants were added together with a picture to the familiarization phase of Experiment 4, as the knowledge of these nouns became relevant here. Pictures of two of the character combinations were reported to be difficult. The four sentence contexts with one of these combinations, including the sentence with unclear bias in the L1 adult group, were removed completely from the materials due to other problems, e.g., they included words unknown to some L2 speakers. To improve discrimination between the other pair (gardener/farmer), the picture of one character (gardener) was slightly changed. Since inspection of the sentences with this character combination revealed no obvious problems for the adult L1 controls, they were not excluded from the analysis of the picture selection task.

[41] In the following visual-world eye-tracking experiment, simple comprehension questions were used to control for participants' comprehension of the sentences and their attention. Here, none of the children showed any problems.

8.4 Experiment 4: Methods

8.4.1 Participants

A group of 27 adult German native speakers, none of whom reported being early bilingual, was recruited for Experiment 4. Three participants were subsequently excluded, one who reported being under medication and two others due to bad calibration and extensive trackloss. Thus, data of 24 adult German native speakers (20 female, 5 left-handed) were included in the final analyses. All of them participated in the experiment for course credit or a payment of six euros. All learners tested in Experiment 4 previously took part in Pre-test 2, which was conducted to evaluate the sentence and picture materials and to establish that the learner groups were sensitive to IC information offline (see section 8.3). For Experiment 4, a group of 24 German speaking children was tested. Two children had to be excluded due to bad calibration and a great deal of movement. Thus, the final analyses included data of 22 children (10 female, 2 left-handed).[42] All of them received a monetary reward of eight euros and a sticker. Twenty-five Russian L1-German L2 speakers participated in Experiment 4. Two participants were excluded, one due to bad calibration and the other one due to not following the instructions. Thus, the final analyses included data of 23 L2 speakers (20 female, 4 left-handed), none of whom reported being early bilingual. All of them learned German at or after the age of seven and were highly proficient in German as measured by the Goethe placement test (Goethe-Institut, 2011). The L2 speakers received a monetary compensation of eight euros for their participation. Table 8.6 below shows the details of the participants included in the analyses of Experiment 4.

The participants' vision was normal or, when necessary, corrected with either glasses or lenses and none reported any speech or hearing disorders. All gave informed written consent regarding the data use; in the case of the children this was done by their caretakers. All procedures were in accordance with the Declaration of Helsinki. Ethical approval for experiments with children for the Potsdam Research Institute for Multilingualism was received from the ethics committee of the University of Potsdam (37/2011). The testing took place in Potsdam and Berlin.

8.4.2 Design & materials

Before the actual experiment, two pre-tests were conducted. First, a forced choice task with adult L1 speakers of German tested the bias strength of a set of German psych verbs (see section 8.2). Verbs with a strong IC bias (80% or more NP1 or NP2 choices) were then taken to construct the sentence materials. These materials were then evaluated in a second pre-test, which used a picture selection task and also included the two learner groups tested in Experiment 4 (see section 8.3). Based on the results

[42] To calculate the age of the children, their exact ages (year; months) were used.

Adult L1 group (n = 24)			
	Mean	SD	Range
Age	24.75	4.68	20–41
Child L1 group (n = 22)			
	Mean	SD	Range
Age	11;8	1;5	9;2–13;9
Adult L2 group (n = 23)			
	Mean	SD	Range
Age	27.96	4.02	21–37
AoA German	12.87	5.38	07–22
Years of exposure to German	15.09	5.38	06–25
Years of immersion	08.52	5.77	01–21
Goethe test	26.57	2.17	22–30

Table 8.6: Participant details for Experiment 4, showing the means, standard deviations and ranges

and observations of this second pre-test, the materials were modified accordingly. For example, the order of characters was fully counterbalanced.

In Table 8.7, an example of an experimental item pair in Experiment 4 is given; the corresponding visual display is shown in Figure 8.4b. The final eye-tracking experiment included 44 experimental and 44 filler sentences from prior pre-testing. Half of the experimental items included an NP1-biasing verb and the other half an NP2-biasing verb. In addition, a second version of each sentence was recorded in which the order of the characters included was reversed, so there were 88 experimental sentences in total distributed across two lists using a Latin square design. Character position served as a control condition and was manipulated within-items; verb bias was the critical condition and a between-items factor. Each list was split into two blocks of 44 trials. The order within blocks was pseudorandomized. Since one character combination used in the picture selection task was excluded here, there were 11 different combinations of male Playmobil® characters for the experimental items. In each block the same character combination appeared twice but in different contexts. Both lists were presented in two different orderings by switching the two blocks. The filler sentences varied in sentence structure, also included female characters and plurals and also used connectives other than *weil* (because). For a full list of the experimental sentences and fillers, see section C.1 in the appendix.

The experimental sentences were constructed in the following way: The distractor, consisting of an adverbial phrase, served the purpose of attracting the eye-movements away from the characters before the causal connective followed by the pronoun was encountered, as in Cozijn et al. (2011). To exclude the

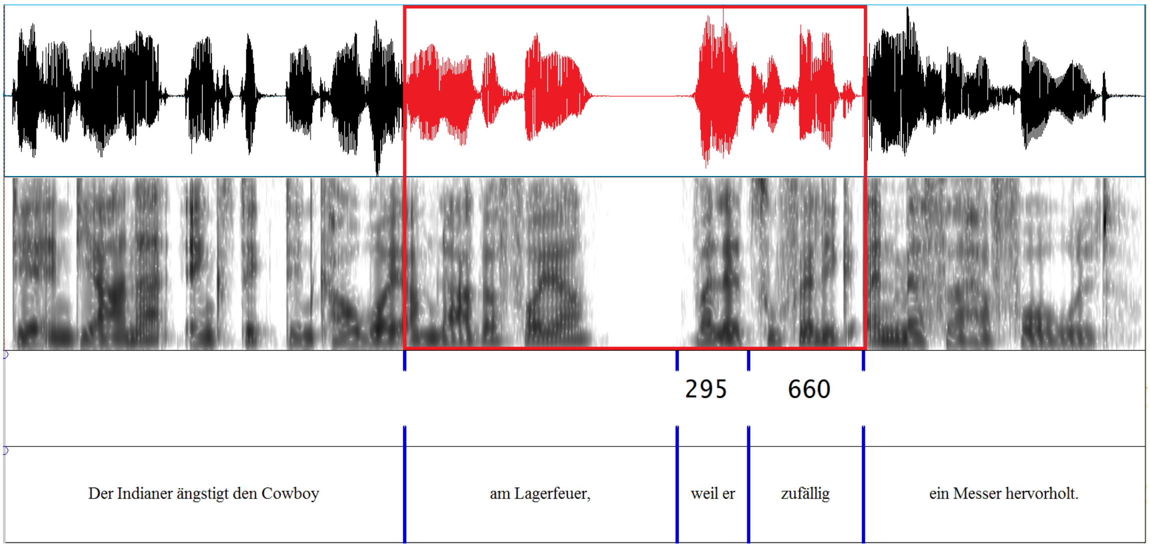
		distractor	CRITICAL WINDOW	disambiguating segment
item	NP1 verb NP2	adverbial	con + pro adverbial	object verb
1	**Der Indianer** ängstigt den Cowboy	am Lagerfeuer,	weil **er** zufällig	ein Messer hervorholt.
	The Native American frightens the cowboy	at the campfire	because **he** accidentally	a knife takes out
	'The Native American frightens the cowboy at the campfire because he accidentally takes out a knife.'			
2	Der Indianer fürchtet **den Cowboy**	am Lagerfeuer,	weil **er** zufällig	ein Gewehr hervorholt.
	The Native American fears **the cowboy**	at the campfire	because **he** accidentally	a rifle takes out
	'The Native American fears the cowboy at the campfire because he accidentally takes out a rifle.'			

Table 8.7: Auditory stimuli in Experiment 4. The critical window for an anticipatory effect is shaded in grey. The referent of the pronoun is in bold.

distractor as a possible referent of the pronoun, only feminine or neuter gender nouns or plurals were used. The critical window included the causal connective *weil* (because), the pronoun *er* (he) and an adverbial with a length of three to four syllables to extend the anticipatory time window. The segment including the distractor and the critical window was spliced from one of the sentences to all others with the same context using Praat (Boersma, 2001) as depicted in Figure 8.4a. Thus, the length and prosody of this segment was identical in all conditions. All sentences were spoken by a female native speaker of German at a normal speaking rate. All contexts included disambiguating information at the end of the sentence. Here, an object was mentioned with which the referent of the pronoun appeared in the visual display, like in Arnold et al. (2000) and Arnold et al. (2007). In most cases, these objects belonged to the same object category (e.g., weapon, musical instrument) and/or had the same color or shape to control for salience. The objects moreover had the same grammatical gender in German so, in a strict sense, the sentences were still ambiguous at the determiner *ein* (indefinite article *a*) of the last NP.

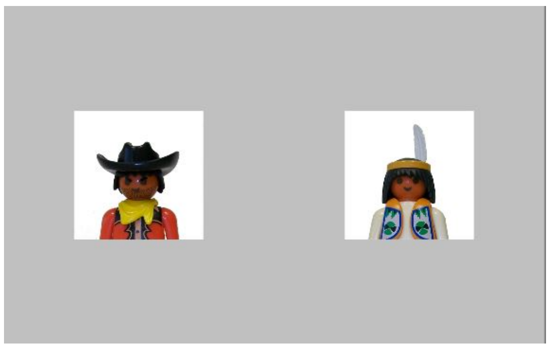
Each sentence was followed by a comprehension question that was spoken by another female native speaker of German. The comprehension questions asked about the actions or states in the sentences (e.g., *Who is taking out a knife?*, *Who is afraid?*), or a location that was mentioned (e.g., *Where does the prince see the princess?*). The comprehension questions were included to ensure that the participants paid attention to the content of the sentences and did not direct their attention towards the pronoun.

The pictures of the characters were photographs of Playmobil® figures and, with one exception, the same ones used in the second pre-test. The distractor pictures were taken from the updated Bank of Standardized Stimuli (BOSS, see Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010) and were modified, or were constructed with the help of other Playmobil® items or toys. All pictures were 400 x 400 pixels with 72 dpi in size; the figures appeared on a white background and the pictures were saved as .jpg-files. The distractor picture was always displayed in the bottom center of the screen. The position of the characters in a list was counterbalanced, so that, overall, the first-mentioned referent appeared



(a) Highlighted in red is the part that was spliced, so it was identical across conditions. The numbers show the average length of the connective + pronoun segment and the adverbial, which constitute the critical time window.

(b) Visual display shown during the sentence



(c) Visual display for the comprehension question

Figure 8.4: Auditory and visual materials in Experiment 4

as often in the top left as in the top right position. The pictures for the comprehension question display remained in the same position as in the prior display when the characters were asked about, however only the upper part of the characters without the mentioned item was shown; see Figure 8.4c.

A cross-linguistic study conducted by Hartshorne et al. (2013) showed an effect of IC in Russian that was not influenced by aspect as it exists in Russian, where verbs can appear as perfective or imperfective. Kim and Grüter (2019) reported cross-linguistic differences that influenced the bias strength for IC verbs in a group of Korean learners of English. However, in a language like Korean, causation can be marked explicitly, which is not the case in Russian. To further check that in the materials used the bias was consistent in German and Russian, two native speakers of Russian translated the verbs into Russian and marked whether the verbs had an NP1- or NP2-bias. Only twice did one of the native speakers display a deviant preference for a verb, which was an NP1-preference for an Experiencer-subject verb; this, however, was not taken as a reason to exclude these verbs, as only one of the two raters ruled out the NP2 reference.

8.4.3 Apparatus

Like in Experiments 1–3, participants' eye-movements in Experiment 4 were tracked with an SMI RED eye-tracker at a sampling rate of 120 Hz. The participants sat at a distance of approximately 65 cm in front of the stimulus screen, which had a resolution of 1680 x 1050 pixels. Both eyes were tracked and their movements averaged. A 9-point calibration procedure was used. The goal was to have a visual acuity below 0.5 degrees. Since no chin rest was used and viewing was not constrained, for the children in particular this criterion could not always be met. Since the picture size was quite big, calibration was sometimes accepted when it was below 0.99 degrees and eye-tracking looked otherwise stable.

8.4.4 Procedure

8.4.4.1 Experimental procedure

First, participants were familiarized with the Playmobil® figures. Then they received the written instructions (oral for the children) for the experiment (see subsection C.1.3 in Appendix C). After calibration/validation, the experiment started with four practice trials. Each trial included a preview of 1000 ms, before the participants listened to the sound file. The visual display remained on the screen for a further 800 ms until the comprehension question followed, together with two pictures displayed on the left and right side of the screen. The participants had to press either the left or the right button on a gamepad to indicate which picture answers the question. The number of left and right correct responses was counterbalanced. The participants were instructed to answer these questions as accurately as possible. They continued to the next trial by pressing a button, giving them the opportunity to do the experiment at their own pace. There was a break after the first block with new calibration afterwards. Whenever necessary, calibration was repeated during the experiment. This part of the experiment lasted between 20 and 30 minutes.

The child L1 and adult L2 groups further completed a forward and backward digit span task, which was adapted from the Hamburg-Wechsler Intelligence Test (Tewes, 1991). To avoid a disadvantage in the adult L2 group, the digits were not presented aurally but rather visually on a laptop screen, where they appeared at intervals of one second. Participants had to write down the digit span they had just seen when a paper and pencil symbol appeared on the screen. The test started with two three-digit spans in the forward subtask and two two-digit spans in the backward subtask, and one digit was added with every row on the score sheet. The test ended when a participant could not recall both digit spans in a row. Participants could encounter 28 trials in total, so participants could reach 28 points maximally, 14 in each subtask.

8.4.4.2 Data pre-processing

All items that included the character combination gardener/farmer were excluded from further analyses, because participants still reported that the pictures were difficult to discriminate (9% of the data). All trials with a verb that an L2 speaker indicated as unknown were excluded on a by-participant basis,[43] as were eight trials with an interruption (1% of the data). Participants in Experiment 4 could fixate one of three AOIs or the background of the screen. Background looks were treated as trackloss (within the critical time window 2% overall). It was expected that participants would fixate the location mentioned before the connective and pronoun segment, here termed the *distractor*, before shifting their attention to the two referent AOIs. Figure 8.5 and Figure 8.6 show an overview of the time course including all AOIs per group and verb bias condition. Except for the adult L2 group, fixation proportions at the onset of the connective and pronoun segment were around chance level (0.33). The L2 speakers, particularly for the NP2-biasing verbs, showed fewer distractor fixations. Crucially, there is no visible difference between the proportion of looks for the NP1 and NP2 referent AOI at the onset of the connective and pronoun segment; only afterwards do the proportions visibly increase as a function of verb bias in the child and adult L1 group and in the adult L2 group for the NP1- but not NP2-biasing verbs.

For the analyses, only the two referent AOIs were included. The critical window for an anticipatory effect started 200 ms after the onset of the connective and pronoun segment and ended 200 ms after the offset of the adverbial prior to the onset of the disambiguating segment, to take into account eye-movement latency. Its length after exclusion of all trials without any fixations on one of the referent AOIs within this time region was 955.5 ms (SD: 66.44 ms). The exact on- and offsets determined in Praat (Boersma, 2001) were used and data points aggregated into 50 ms time bins. In the adult L1 group, 90% of the data included fixations on the NP1 or NP2 referent, in the child L1 group 92% and in the adult L2 group 96%. For all analyses and the described pre-processing, the eye-trackingR package (Dink & Ferguson, 2015) was used.

[43] Here, the information from the second pre-test was used, where participants received a list of the verbs and some of the nouns to indicate whether they were familiar with them. It was not assumed that participants had learned these words through exposure in that experiment, which was conducted a while ago.

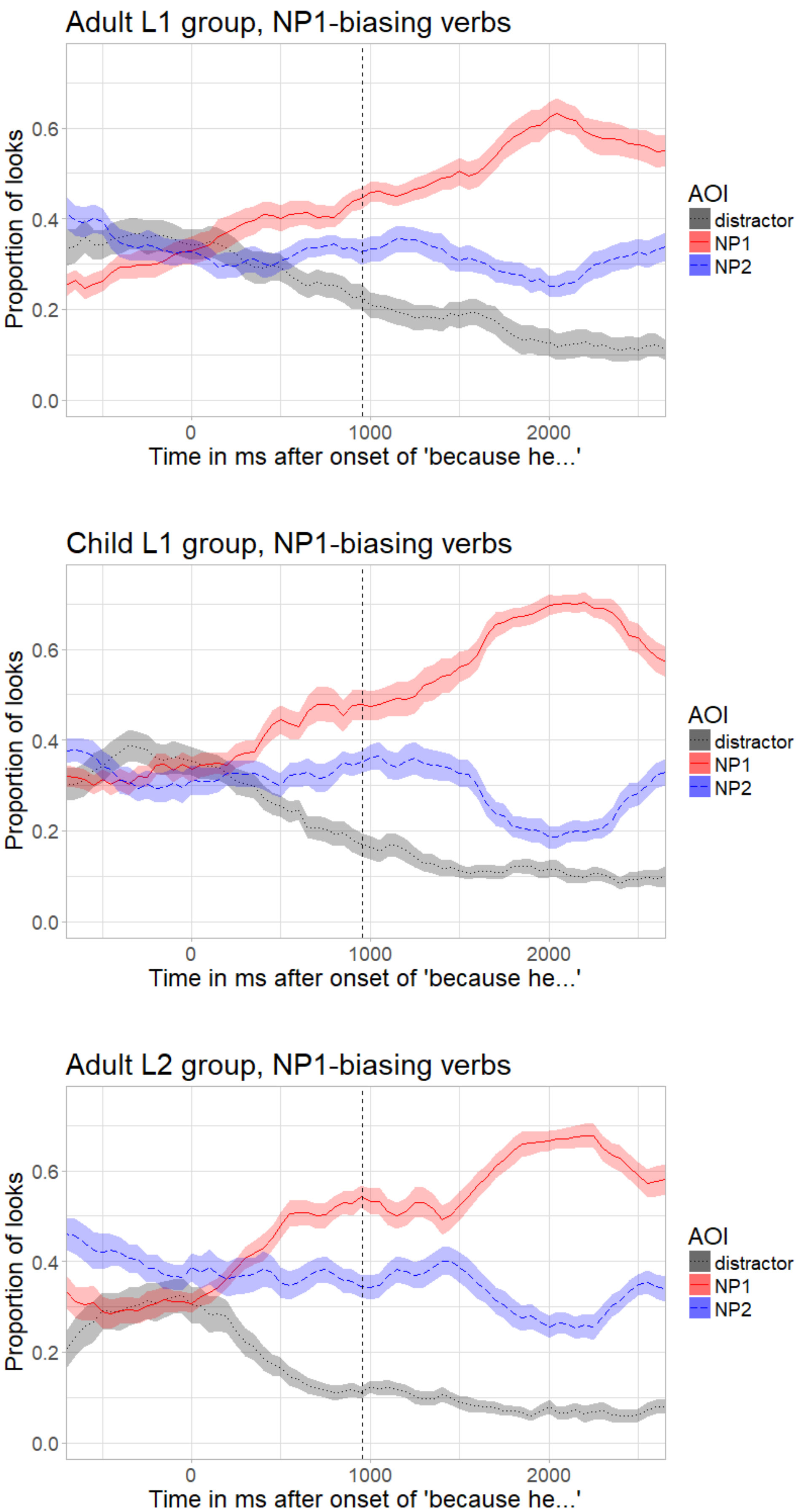


Figure 8.5: Proportion of looks for all AOIs for the NP1-biasing verbs in Experiment 4 for the adult L1 (top), child L1 (middle) and adult L2 groups (bottom). The connective onset shifted 200 ms forwards is aligned to zero; the dashed vertical line marks the mean onset of the disambiguating information.

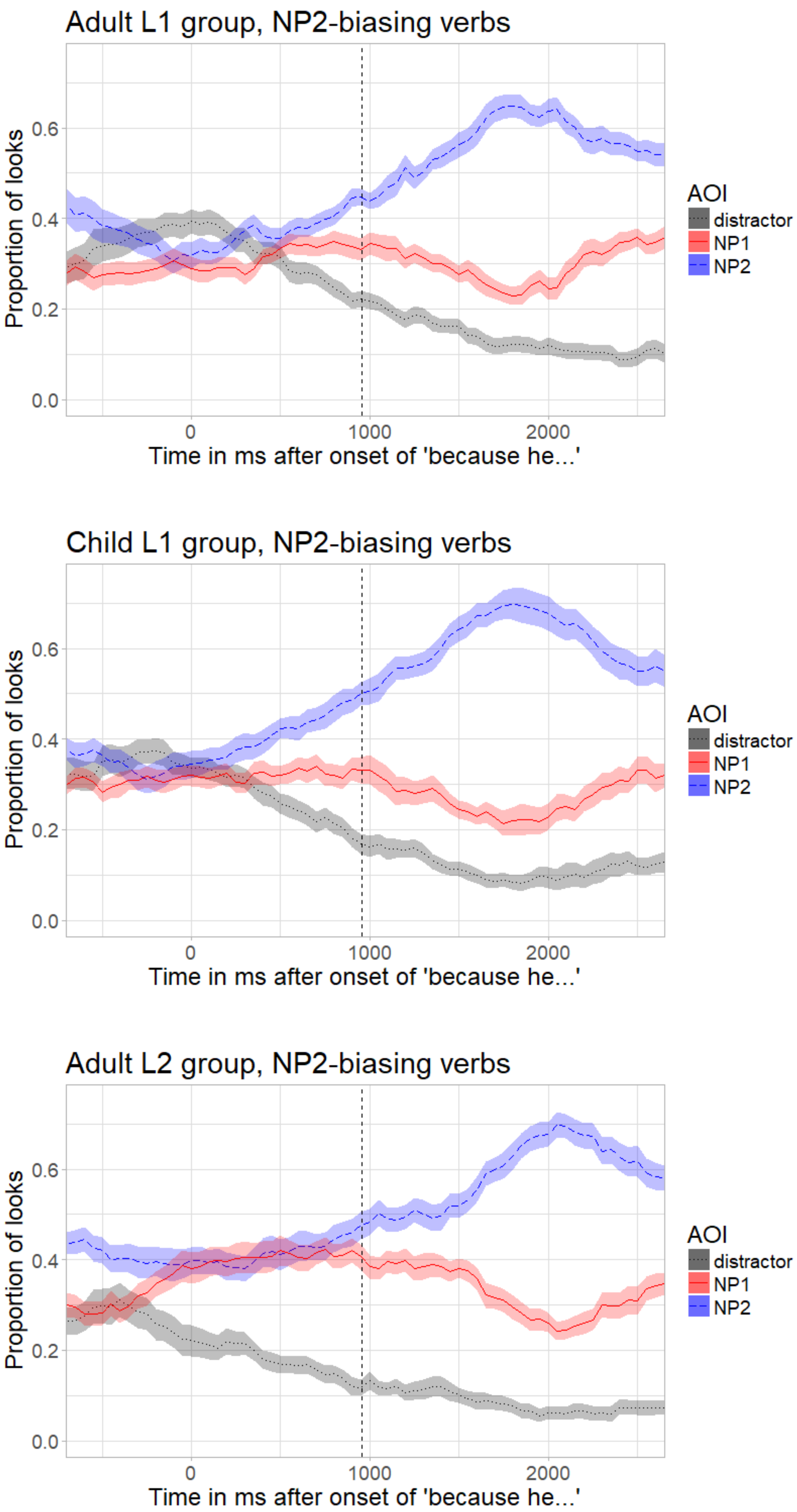


Figure 8.6: Proportion of looks for all AOIs for the NP2-biasing verbs in Experiment 4 for the adult L1 (top), child L1 (middle) and adult L2 groups (bottom). The connective onset shifted 200 ms forwards is aligned to zero; the dashed vertical line marks the mean onset of the disambiguating information.

8.5 Results

8.5.1 Behavioral data

The adult L1 group had a response accuracy of 98% on average, ranging from 93% to 100% (SD: 2%). The child L1 group answered on average 95% of the comprehension questions correctly, ranging from 86% to 99% (SD: 4%), and the adult L2 group answered 93% correctly with a range from 83% to 100% (SD: 6%). The behavioral data demonstrate that the participants in all groups had no difficulties in understanding the sentence contexts they were presented with and that they paid attention.

8.5.2 Digit span task

Both learner groups also completed a digit span task to assess their memory capacity and confirm the assumption that the children were cognitively less mature. In the forward subtask, the children produced 5.59 correct trials on average (SD: 1.99, range: 3–10) and the adult L2 speakers 7.96 (SD: 2.53, range: 4–13). In the backward subtask, correctly recalled trials in the child group amounted to a mean of 6.41 (SD: 2.24, range: 3–11) and in the adult L2 group to a mean of 9.57 (SD: 2.27, range: 5–14). A simple logistic regression confirmed that the adult L2 speakers scored better than the children in the forward subtask as well as in the backward subtask; see Table 8.8 for the model output and formulas. Thus, the digit span task demonstrated that the children as a group indeed had a lower memory capacity than the adult L2 speakers as a group. The backward task in particular is associated with working memory, as here the digits are not only recalled but also have to be stored and manipulated (see Huettig & Janse, 2016).

Forward subtask	Estimate	Std. Error	*z*-value	*p*-value
Intercept (Group = Child L1)	-0.4082	0.1163	-3.508	0.0005 ***
Adult L2	0.6832	0.1618	4.221	< 0.0001 ***
Formula: glm(cbind(learnerWM$forward, 14 - learnerWM$forward) ~ learnerWM$group, family = binomial)				
Backward subtask	Estimate	Std. Error	*z*-value	*p*-value
Intercept (Group = Child L1)	-0.1692	0.1144	-1.480	0.139
Adult L2	0.9379	0.1656	5.663	< 0.0001 ***
Formula: glm(cbind(learnerWM$backward, 14 - learnerWM$backward) ~ learnerWM$group, family = binomial)				
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

Table 8.8: Results of the analyses of the digit span task

8.5.3 Eye-tracking data

As mentioned in the introduction of this chapter, personal pronouns typically refer to the first-mentioned entity and/or subject of the matrix clause. Hence, if IC information is used as a predictive cue, the likelihood that the NP1 referent is fixated should be different for the two verb bias conditions, with a reduced likelihood for NP2-biasing verbs. Therefore, the following linear mixed-effects models analyzed the Elog-transformed fixation proportions for the NP1 referent. I will start with separate models for each group, before presenting the group analysis. Here, the fixed effects comprised the between-items factor bias (NP1 vs. NP2), linear time and their interaction. The random effects comprised subjects and items as random intercepts, as well as the interaction between bias and linear time as by-subject slope and linear time as by-item slope, which corresponds to the maximal model (Barr et al., 2013).

The first model, see Table 8.9 for the output and Figure 8.7 for the visualization, shows the results for the reference group of adult L1 speakers. Treatment contrast coding was used for verb bias, so the intercept corresponds to the likelihood that the NP1 referent was fixated within the critical time window for the NP1-bias condition. The effect of verb bias demonstrates that, overall, adult L1 speakers of German were less likely to fixate the NP1 referent if the verb biased towards the NP2 referent, demonstrating an early effect of IC information and anticipation of the target referent.

Adult L1 group	Estimate	Std. Error	*t*-value	*p*-value
Intercept (bias = NP1)	0.2445	0.1593	1.535	0.1310
Bias (NP2)	-0.4626	0.1907	-2.426	0.0186 *
Linear Time	0.2873	0.5784	0.497	0.6216
Bias x Linear time	-0.2394	0.7614	-0.314	0.7545
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Formula: Elog ~ bias*(ot1) + (1 + (ot1) | item) + (1 + bias*(ot1) | subject)				

Table 8.9: Results of the growth curve analysis for the adult L1 group in Experiment 4

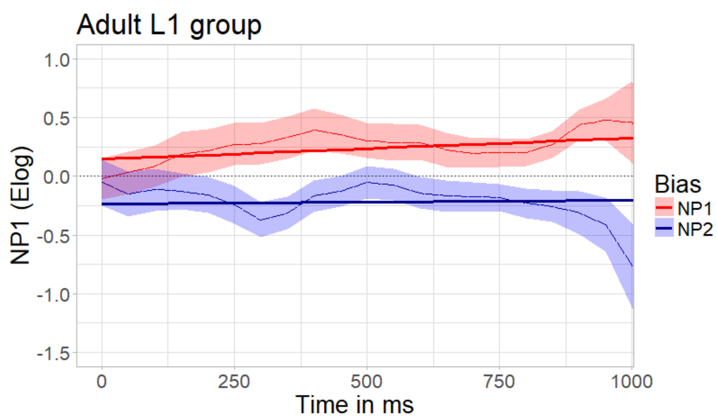


Figure 8.7: Time course showing the probability of looks to the NP1 referent during the critical window in Experiment 4 for the adult L1 group; curves show the raw data (light) and model prediction (bold)

Table 8.10 shows the model output for the child L1 group. Like the adult L1 speakers, the children were less likely to fixate the NP1 referent for NP2-biasing verbs. A marginal interaction between verb bias and linear time points to a different development for the two verb bias conditions. When taking the NP2-bias condition as the baseline, which was done by re-leveling the factor verb bias, there was no significant effect of linear time. Instead, a significant effect at the intercept indicates that the likelihood that the NP1 referent was fixated was, on average, below chance level in the child L1 group.

Child L1 group	Estimate	Std. Error	*t*-value	*p*-value
Intercept (bias = NP1)	0.2768	0.1795	1.542	0.1314
Bias (NP2)	-0.6007	0.2646	-2.270	0.0293 *
Linear time	0.5097	0.4953	1.029	0.3088
Bias x Linear time	**-1.2481**	**0.6946**	**-1.797**	**0.0789 .**
Intercept (bias = NP2)	-0.3239	0.1472	-2.201	0.0327 *
Linear time	-0.7384	0.5431	-1.360	0.1807
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				
Formula: Elog ~ bias*(ot1) + (1 + (ot1) | item) + (1 + bias*(ot1) | subject)				

Table 8.10: Results of the growth curve analysis for the child L1 group in Experiment 4. Below the results of the model with NP1-bias as baseline, the model with NP2-bias as baseline is shown to follow up on the interaction (highlighted).

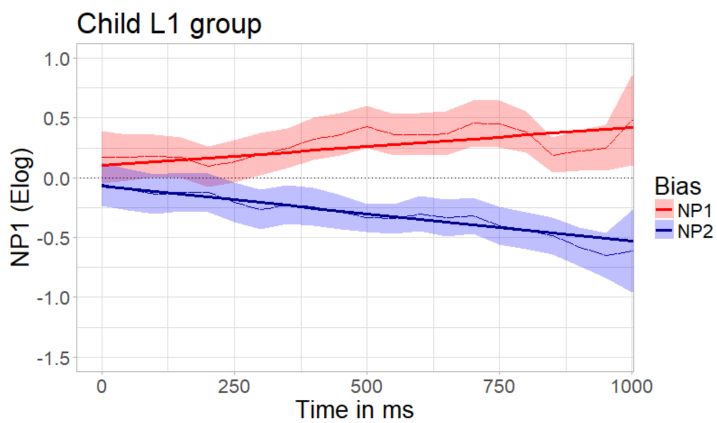


Figure 8.8: Time course showing the probability of looks to the NP1 referent during the critical window in Experiment 4 for the child L1 group; curves show the raw data (light) and model prediction (bold)

The results of the model for the adult L2 group, see Table 8.11, show that an effect of verb bias is absent. Instead, the likelihood that the NP1 referent was fixated for the NP1-biasing verbs increased linearly, as indicated by the effect of linear time. The corresponding graph in Figure 8.9 shows that the probability of looks to the NP1 referent for NP2-biasing verbs remained stable around zero, i.e., around chance level, even towards the end of the critical window. After re-leveling the factor verb bias, the intercept had an estimate of 0.0059.

Adult L2 group	Estimate	Std. Error	*t*-value	*p*-value
Intercept (bias = NP1)	0.2868	0.1941	1.478	0.1461
Bias (NP2)	-0.2809	0.2471	-1.137	0.2606
Linear time	0.9952	0.4409	2.257	0.0289 *
Bias x Linear time	-1.0515	0.7461	-1.409	0.1657
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				
Formula: Elog ~ bias*(ot1) + (1 + (ot1) | item) + (1 + bias*(ot1) | subject)				

Table 8.11: Results of the growth curve analysis for the adult L2 group in Experiment 4

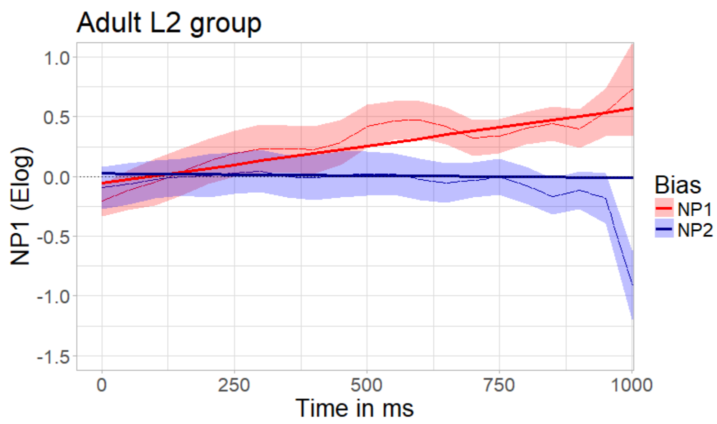


Figure 8.9: Time course showing the probability of looks to the NP1 referent during the critical window in Experiment 4 for the adult L2 group; curves show the raw data (light) and model prediction (bold)

The separate analyses showed an effect of IC in terms of a reduced likelihood that the NP1 referent was fixated in the NP2-bias condition for the adult and child L1 group but not for the adult L2 group. In contrast, the adult L2 group showed an increasing likelihood that the NP1 referent was fixated for NP1-biasing verbs. There was further indication that the children were even less likely to fixate the NP1 referent for NP2-biasing verbs. Next, all groups were analyzed together. Table 8.12 shows the output of the maximal model using a priori contrasts (Schad, Hohenstein, Vasishth, & Kliegl, July 2018), which tested the effect of verb bias and linear time across groups but taking one group as reference, so effects of and interaction with group can be obtained. As indicated by the effect of verb bias, the anticipatory effect showed up across groups. A marginal effect of linear time indicates a linear increase in the likelihood that the NP1 referent was fixated for NP1-biasing verbs, and the marginal interaction between verb bias and linear time indicates that the slope was different for NP2-biasing verbs. A significant effect at the intercept indicates that across groups the likelihood that the NP1 referent was fixated was above chance level for NP1-biasing verbs. The model with the adult L1 group as reference shows the comparison between L1 groups and the adult groups; the model with the child L1 group as reference shows the comparison between learner groups. Although each group showed a slightly different pattern, as seen in the separate analyses above, no significant effects of group emerged in the group analyses. It seems likely that the differences were just too subtle to lead to an effect of or interaction with the factor group.

Reference group: Adult L1	Estimate	Std. Error	*t*-value	*p*-value
Intercept (NP1-bias, across groups)	0.2687	0.104	2.585	0.0112 *
Bias (NP2, across groups)	-0.4469	0.1373	-3.255	0.0016 **
Child L1	0.0325	0.2389	0.136	0.892
Adult L2	0.041	0.257	0.159	0.8737
Linear time (across groups)	0.5924	0.3187	1.859	0.0675 .
Bias x Child L1	-0.1368	0.3123	-0.438	0.6623
Bias x Adult L2	0.1816	0.3403	0.533	0.595
Bias x Linear time (across groups)	-0.8406	0.4583	-1.834	0.0709 .
Child L1 x Linear time	0.2273	0.7293	0.312	0.7562
Adult L2 x Linear time	0.7241	0.7057	1.026	0.3083
Bias x Child L1 x Linear time	-1.0012	1.0519	-0.952	0.3443
Bias x Adult L2 x Linear time	-0.8369	1.0178	-0.822	0.4136
Reference group: Child L1	Estimate	Std. Error	*t*-value	*p*-value
Adult L2	0.0085	0.2529	0.033	0.9734
Bias x Adult L2	0.3184	0.333	0.956	0.3415
Adult L2 x Linear time	0.4968	0.6372	0.780	0.4379
Bias x Adult L2 x Linear time	0.1643	0.9236	0.178	0.8593
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Formula: Elog ~ bias*group*(ot1) + (1 + group*(ot1) | item) + (1 + bias*(ot1) | subject)				

Table 8.12: Results of the group analyses for the critical window in Experiment 4

A post-hoc analysis of the adult L1 data took into account that the model in Table 8.9 did not actually provide a good fit, as shown by the regression lines in Figure 8.7. Note that above only linear time was included, as there was no hypothesis as to why the time course should show a non-linear development. It was expected that at the onset of the connective and pronoun segment participants would start shifting their gaze to the referent AOI that is congruent with the verb bias. In a post-hoc analysis, higher order polynomials were added to the model shown in Table 8.9. Since the maximal model with by-item and by-subject slopes for all time terms did not converge, the model was simplified and the one with the lower AIC value taken. This analysis demonstrated an effect of cubic time for the adult L1 speakers in accordance with the shape of the curves, see Figure 8.10.[44] A possible explanation for the non-linear development is that the adult L1 speakers, at least temporarily, considered the other character as a potential referent. Indeed, previous research with adults has reported such a contrast effect (Kamide, Altmann, & Haywood, 2003; Sedivy, Tanenhaus, Chambers, & Carlson, 1999): Although looks to the

[44] Including higher order polynomials up to the third term did not lead to an effect of quadratic or cubic time in the other two groups.

target picture had already started to increase, still more looks to the contrasting object relative to unrelated distractors were observed that could indicate a process of ‘double-checking.’ The reason why this process mainly affected the NP2-biasing verbs could be an overall NP1-preference common in studies on pronoun resolution. Note that in this ‘improved’ model the probability of looks to the NP1 referent was, on average, below chance level in the NP2-bias condition, like for the child L1 group before, and close to being significantly above chance level for NP1-biasing verbs.

Adult L1 group	Estimate	Std. Error	*t*-value	*p*-value
Intercept (bias = NP1)	0.2557	0.1456	1.756	0.0853 .
Bias (NP2)	-0.5595	0.1767	-3.167	0.0026 **
Linear time	0.413	0.5804	0.712	0.4806
Quadratic time	-0.1302	0.4018	-0.324	0.7478
Cubic time	0.371	0.2154	1.722	0.0884 .
Bias x Linear time	-0.9981	0.7808	-1.278	0.2078
Bias Quadratic time	-0.5359	0.5713	-0.938	0.3543
Bias x Cubic time	**-0.9705**	**0.3090**	**-3.140**	**0.0022 ****
Intercept (bias = NP2)	-0.3038	0.1359	-2.235	0.0297 *
Linear time	-0.5851	0.5984	-0.978	0.3338
Quadratic time	-0.6661	0.4429	-1.504	0.1408
Cubic time	-0.5996	0.2217	-2.704	0.008 **

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Formula: Elog ~ bias*(ot1+ot2+ot3) + (1 + (ot1+ot2+ot3) | item) + (1 + bias*(ot1+ot2) | subject)

Table 8.13: Results of the growth curve analysis for the adult L1 group in Experiment 4 (post-hoc). Below the results of the model with NP1-bias as baseline, the model with NP2-bias as baseline is shown to follow up on the interaction (highlighted).

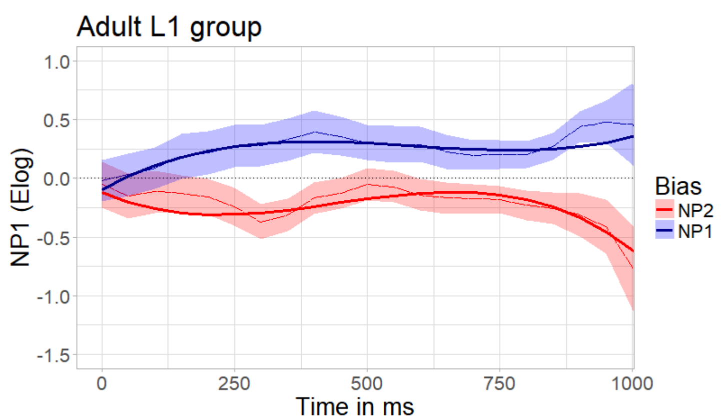


Figure 8.10: Time course showing the probability of looks to the NP1 referent during the critical window in Experiment 4 for the adult L1 group (post-hoc). The curves show the raw data (light) and model prediction (bold).

Since the memory capacity of the child L1 and adult L2 groups had also been tested, it appeared informative to include it as a predictor variable to see whether it mediates prediction. Separate analyses for the child L1 and adult L2 group were conducted that included the centered values of the forward and the backward subtask. For both groups, the maximal models did not show a significant effect of or interaction with the values in either the forward or the backward subtask. Hence, there is no indication that in the current experiment memory capacity influenced the results within groups.

Next, the onset of the anticipatory effect was determined for each group. Note that the separate growth curve analyses revealed an overall effect of the IC bias for the L1 groups only, whereas for the adult L2 group IC information only became effective over time. A bootstrapped cluster-based permutation analysis (Maris & Oostenveld, 2007) as implemented in the eyetrackingR package (Dink & Ferguson, 2015) was conducted separately for each group tested in Experiment 4. Here, the dependent measure corresponds to the NP1 referent fixation proportion within each 50 ms time bin in the critical window, when taking into account only the two referent AOIs. This analysis tests in which time bins the curves significantly diverged, and the fixation proportions for NP1- and NP2-biasing verbs were significantly different from each other. For better illustration, the graphs in Figure 8.11 show the proportion of looks for the NP1 referent during, but also before and after the critical window to also show potential baseline differences and the difference between conditions when the pronoun is disambiguated.

A paired t-test was run on each time bin in the critical time window quantifying the statistical significance of the effect, thus identifying time bins where conditions significantly differ. Adjacent time bins that passed the threshold statistics (α-level = 0.05, two-tailed) were then grouped together into clusters. In a next step, a null distribution was bootstrapped by shuffling the data in each time cluster (here: 1000 times). The observed data were then compared to this bootstrapped distribution. Table 8.14 shows the clusters with a difference between the two verb bias conditions. The length of the critical window in Experiment 4 varied between items (min: 854 ms, max: 1110 ms), therefore the last time bins included a smaller number of distinct samples, which should be considered for the end time.

Group	Cluster	Direction	SumStatistic	StartTime	EndTime	Probability
Adult L1	1	Positive	52.687	250	1150	0.000
Child L1	1	Positive	40.913	400	1150	0.008
Adult L2	1	Positive	4.873	600	700	0.151
	2	Positive	16.834	800	1100	0.037

Table 8.14: Results of the cluster-based permutation analysis in Experiment 4. The table shows the start and end times for the clusters identified and the probability of seeing the effect by chance.

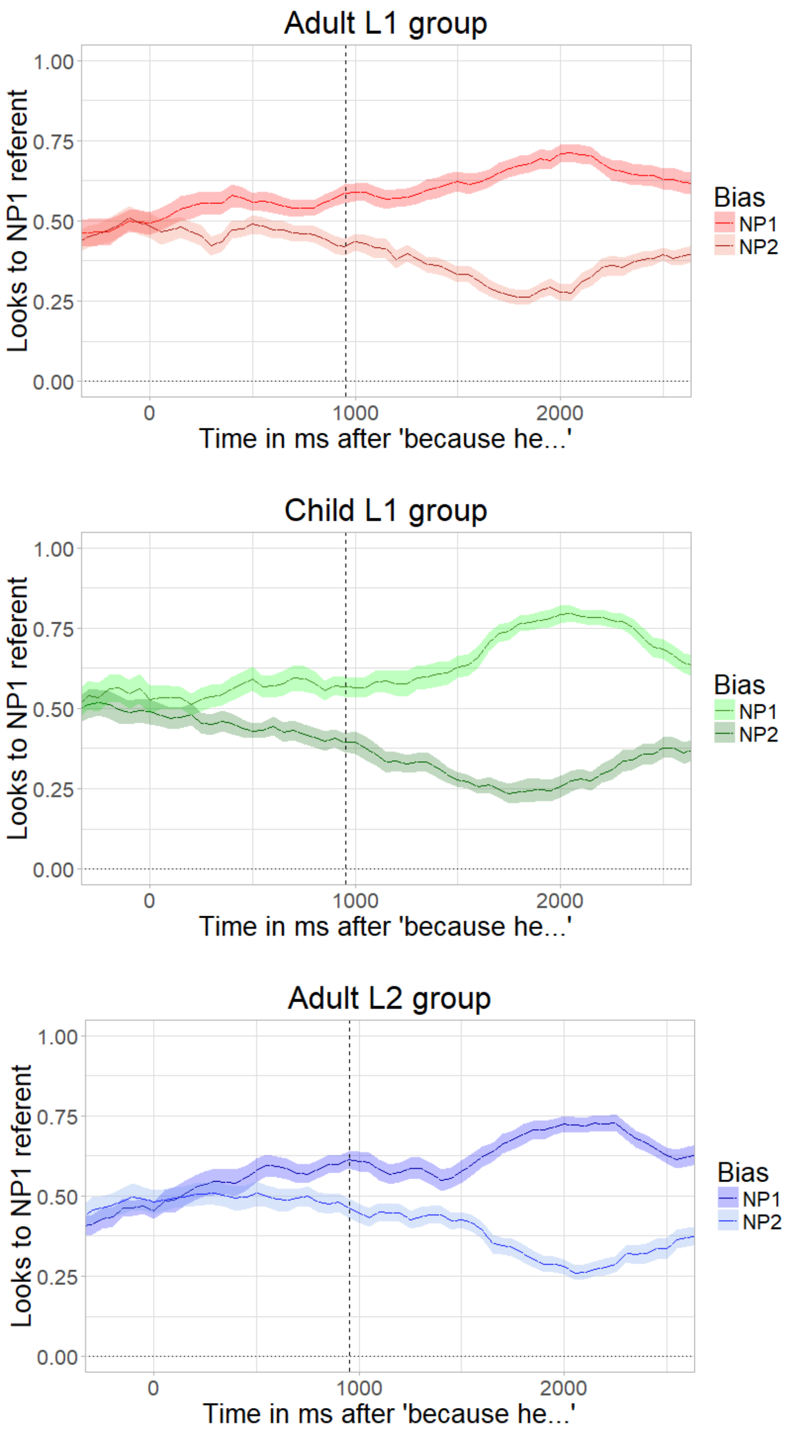


Figure 8.11: Time course showing the proportion of looks to the NP1 referent before, during and after the critical window in Experiment 4 for the adult L1 (top), child L1 (middle) and adult L2 groups (bottom). The connective onset shifted 200 ms forwards is aligned to zero; the dashed vertical line marks the mean onset of the disambiguating information.

The analyses show that the onset of the anticipatory effect was at around 250 ms in the adult L1 group and at around 400 ms in the child L1 group. Thus, the cluster-based permutation analysis reveals a delay of around 150 ms for the children. Even later, only at around 800 ms after the onset of the critical window did the proportion of looks to the NP1 referent significantly diverge between the two bias conditions in the adult L2 group, which is line with the results of the separate growth curve analysis. For the adult L1 control group, the effect emerged within the connective and pronoun segment, which had an average length of 295 ms, which is in line with previous studies on the use of IC information in L1 speakers.

8.6 Discussion

Experiment 4 investigated the predictive use of IC information and, in addition to an adult L1 group, included two different learner groups of German, a child L1 and an adult L2 group. Three pictures were presented to the participants, the two characters mentioned and a location, whereby the latter only had the function of a gaze neutralizer and was not included in the analyses. Importantly, the results for the adult native speakers of German could replicate those from previous studies on the use of IC information in Finnish and Dutch (Cozijn et al., 2011; Järvikivi et al., 2017; Pyykkönen & Järvikivi, 2010). Shortly after the causal connective and clearly before the disambiguating information was encountered, here provided by an object the characters appeared with, the adult L1 control group was less likely to fixate the NP1 referent for NP2-biasing verbs, demonstrating anticipation of the biased antecedent and, thus, prediction of upcoming reference.

This experiment compared for the first time the use of IC information in different learner groups to that of an adult L1 control group, while using the same experimental design. The first research question addressed was whether child L1 and adult L2 speakers of German use IC information in the same way as adult L1 speakers to predict upcoming reference. Although the group analysis did not show an effect of group or any interactions with group, subtle differences were seen when analyzing the groups separately: Unlike the adult L1 and child L1 group, the adult L2 group showed no effect of condition, meaning no difference between verb bias conditions for the average critical time window. However, an effect of linear time in the L2 group indicated an increase in looks to the target referent for NP1-biasing verbs. Therefore, the first research question can be answered positively in the case of the child group but only with reservations in the case of the adult L2 group. The second research question concerned the onset of the anticipatory effect. For the adult L1 control group, the NP1 fixation proportions differed between verb bias conditions 250 ms after the corrected onset of the critical window, that is, within the connective and pronoun segment. The children were slightly delayed (around 150 ms later). The adult L2 group, in contrast, showed an onset of divergence only after 800 ms, that is, towards the end of the critical window. Hence, the results point to a reduced ability to generate expectations at the discourse-level as proposed by the RAGE hypothesis, here in terms of a later onset and a different pattern for NP1- and NP2-biasing verbs. Below, the results for each group will be discussed in more detail.

The results of Experiment 4 add to existing research on the predictive use of IC information in adult native speakers by finding an effect in German, further supporting the assumption that it is a cross-linguistically consistent phenomenon (Hartshorne et al., 2013). Interestingly, the adult L1 group displayed a non-linear development, which was interpreted in terms of a process of 'double-checking.' Although the NP1 fixation proportions started to diverge around 250 ms after the onset of the critical window, the L1 speakers again seemed to consider the other referent as a potential antecedent of the ambiguous pronoun. The growth curve analysis testing the likelihood that the NP1 referent was fixated, including also quadratic and cubic time, showed that this process was enhanced for the NP2-biasing verbs, probably as a result of an NP1-preference. Such a preference had also been observed in other studies and, as discussed in the introduction, is also known as a first-mention advantage. In the study by Järvikivi et al. (2017) it also became effective after the IC information. In Experiment 4, the NP1 referent was always also the subject of the matrix clause.

Experiment 4, for the first time, showed that German speaking children between the ages of nine and 14 years integrated multiple information sources, here the argument structure of the verb and discourse structure, to predict upcoming reference. Nevertheless, the children were still slower than the adult L1 control group. An additional digit span task confirmed that the children as a group had a lower memory capacity than the adult L2 group tested. Interestingly, the separate analyses showed the same overall effect of IC information in the L1 groups but not in the L2 group. Hence, resource deficits are less likely to account for reduced prediction at the discourse-level based on IC information. Since in the current experiment information about the referent who is the causer of the emotion described had to be retrieved, memory capacity was considered a critical factor. Considering the quote by Grüter et al. (2014) under subsection 4.2.6, resources should be allocated to coreference establishment due to its importance for discourse comprehension. One may speculate that this is what the children in the current experiment did. The children showed a slightly different development than the adult L1 group, which might be the result of a recency effect that was also reported for children at this age group in a sentence completion task conducted by Goikoetxea et al. (2008) in Spanish: The children quickly fixated the target referent in the NP2-bias condition and, unlike the adult L1 speakers, did not again consider the non-target referent. The reason for this recency effect might in fact be (i) their limited (working) memory capacity, although it seemed not to modulate their prediction, (ii) the fact that the NP2-biasing verbs included very frequent psych verbs like *love* or *like* and were thus 'stronger,' and/or (iii) the fact that an NP1-preference for pronouns develops with age (e.g., Arnold et al., 2007). Previous experiments testing children's use of multiple information sources reported difficulties in the integration of cues (see subsection 3.1.1). It is possible though that there is a difference between prediction and re-analysis. One should note though that the children in the current experiment were older than the age groups typically tested. Future research should test the predictive use if IC information in younger age groups and/or also include bias-incongruent sentence continuations to test their ability to revise an initial coreference interpretation.

The separate growth curve analysis for the adult L2 group showed an effect of linear time on the likelihood that the NP1 referent was fixated for the NP1-biasing verbs. As in the sentence completion task conducted by Cheng and Almor (2017), the L2 speakers in Experiment 4 demonstrated an overall NP1-preference. If their subject/first-mention preference was supported by the IC bias of the verb, the effect seemed to be 'boosted.'[45] Due to the overall NP1-preference, the time point at which the NP1 fixation proportions significantly diverged was only towards the end of the critical time window. A later effect was also observed in Experiment 1 for the prediction of an upcoming direct object based on the lexical-semantics of the verb. Recently, Contemori and Dussias (2018) also reported a different timing for early L2 speakers, which the authors attributed to a lower quality of the lexical representation of IC verbs as a consequence of less exposure to the L2.

Can lexical access and/or representation explain the distinctive NP1-preference observed in the data? Although observed for adult L1 speakers as well, IC information preceded an effect of order-of-mention in Järvikivi et al. (2017), and for the adult L1 group in the current experiment IC information also became effective soon after the connective and pronoun onset, with the other referent only being considered again later. In contrast, an effect of order-of-mention and/or subjecthood appeared to precede the effect of IC for the L2 speakers in Experiment 4: Whereas target fixations increased for NP1-biasing verbs, they remained at chance level for NP2-biasing verbs. The results can only partly be explained in terms of slower lexical access or a lower quality of the lexical representation and more likely show that information sources were differently weighted, as discussed below. Very recently, Kim and Grüter (2018) presented a study on the predictive use of IC information in Korean learners of English with different proficiency levels. The sentences were embedded in a context and the materials also included bias-incongruent continuations. Only the two characters were displayed. Unlike in the L1 group, the effect of IC information emerged very late in the L2 group, but there was no NP1-preference. One should note though that here participants were still fixating the NP2 referent when the connective and pronoun segment was encountered (*Justin frightened Steve at first sight because he was wearing a ghost costume*). Participants were instructed to click on the target character after a question (*Who was wearing a ghost costume?*). In Experiment 4, the questions did not strategically target the referent of the pronoun.

The L2 speakers tested in Experiment 4 did predict, however they first adopted a first-mention/subject strategy, showing that they may, at least initially, give priority to surface-level cues to pronoun resolution. Hence, if interpreted in terms of slower or weaker prediction due to different cue weighting, then the results are in line with the RAGE hypothesis. The NP1-preference observed in Experiment 4 might result from the experimental design and the use of a personal pronoun. The sentence completion experiments

[45] Interestingly, the NP1 preference was not observed in the previous offline experiment (Pre-test 2) where the same L2 speakers showed significantly fewer NP1 responses for NP1-biasing verbs than the adult L1 control group and a similar number of NP1 responses for both verb types to the children. This finding may indicate that the NP1 preference was only temporary and could be canceled out by IC information later on, when both clauses were integrated to form one interpretation.

reviewed in subsection 4.2.6 found a subject preference after pronoun prompts (Cheng & Almor, 2017; Grüter et al., 2014, 2017). Cheng and Almor (2019), who conducted a story continuation task with and without a pronoun prompt with L1 and L2 speakers, found that, when a pronoun was provided, L2 speakers had a stronger preference for continuing an NP2-biasing sentence context with the NP1 referent than L1 speakers. Personal pronouns most often refer to the first-mentioned entity or subject, in most cases also the Agent in a sentence, adding a semantic component (Järvikivi et al., 2017), and L2 speakers may rely more on semantic and/or probabilistic information, an explanation consistent with the assumptions of the SSH. The results of Experiment 4 can also be interpreted in favor of the IH, as Cheng and Almor (2017) did, as implicit causality clearly is a phenomenon at the interface of syntax, semantics and pragmatics. According to the IH, the allocation of resources rather than resource limits may affect the syntax-pragmatics mapping, which is in line with the different patterns for child L1 and late L2 learners in the current experiment, where the child L1 group despite a lower memory capacity more closely resembled the adult L1 pattern and showed an earlier onset than the adult L2 group. In the current design, the pronoun directly followed the causal connective, which explicitly signaled an Explanation relation. The results indicate that the L2 speakers differed from the L1 speakers in their real-time integration of cues. In a VW experiment on the use of event structure conveyed by grammatical aspect, Grüter et al. (2016) found that the L2 group made use of verb semantics to anticipate the Goal referent after a verb like *bring*, but their prediction was not modulated by grammatical aspect (*was bringing*), unlike in the L1 group. As discussed in the previous chapter, an experiment by Hopp (2015) showed that L2 speakers also relied more on verb semantics than on case marking. More recently, Grüter et al. (2018a) explained their findings for the predictive use of classifiers in a similar way. Whereas the L1 group seemed to first rely on form class with semantics being a secondary cue, the L2 group showed a greater reliance on semantic information.

How do the results fit into the different acquisition situations of child L1 and late L2 learners described under section 3.1? One may assume that the L2 speakers used their knowledge about personal pronouns in subject position to establish coreference with the first-mentioned entity and subject, which in the case of NP2-biasing contexts was in conflict with the verb bias. Hence, target fixations only increased for NP1-biasing verbs. For NP2-biasing verbs, in contrast, L2 speakers waited for upcoming information to appear. As discussed previously, prediction probably facilitates sentence processing but is not a necessary prerequisite for successful sentence interpretation.

8.7 Conclusion

Experiment 4 tested the use of a discourse-level cue in an adult L1, child L1 and adult L2 group. Both L1 groups quickly made use of the implicit causality of the verb in the matrix clause (e.g., *The Native American frightens/fears the cowboy at the campfire*) to predict upcoming reference, although the child group was slightly slower. Before information about the referent was available, they showed a difference between verb bias conditions. The adult L2 group, however, only showed an anticipatory effect for the NP1-biasing verbs. It was concluded that highly proficient L2 speakers could predict upcoming reference but prioritized surface-level cues (first-mention/subjecthood). They were able to use a discourse-level cue like IC information but only did so if it supported their first preference. Hence, the current results support the RAGE hypothesis for prediction at the discourse-level and, moreover, support the more recent assumption that L2 speakers show a different cue weighting. In addition, the comparison between learner groups who differ in their memory capacity allowed a resource deficit explanation to be ruled out and pointed rather to an L1/L2 difference.

Chapter 9

General discussion

9.1 The research questions

The overarching research question followed throughout this thesis was whether L2 speakers' predictive processing abilities were selectively limited. As shown in the literature review in chapter 4, findings from previous studies could not fully support the RAGE hypothesis, which proposes a reduced ability to generate expectations in L2 speakers. Instead, findings from previous studies demonstrated that L2 speakers' predictive abilities were most clearly limited for prediction at the level of phonology, for prediction based on a morphosyntactic cue like case marking and for prediction based on discourse-level cues like event structure and, probably, implicit causality. Here, the L2 speakers showed either no effect of prediction or not the same effect as the L1 group. Previous studies further found that L2 prediction was affected by a range of factors like L2 proficiency and L1-L2 similarity. Since previous studies did not always control for these factors, a missing effect of prediction could indicate a difficulty either with prediction as such or with the predictive cue at hand. A difficulty with a specific predictive cue could result from individual differences due to factors like L2 proficiency and cross-linguistic differences as mentioned above or it could indicate selective differences specific to L2 processing in line with L2 processing accounts that assume more substantial L1/L2 differences. I argued that prediction may interact with the linguistic domain. More concretely, it was assumed that L2 speakers show a less robust use of grammatical information and difficulty in integrating information sources from different linguistic domains. The aim of the current thesis was to better understand predictive processing in late bilinguals and to pursue the question of why their predictive processing abilities appear to be reduced, thus trying to locate the origin of L1/L2 differences in the application of prediction as a processing mechanism.

For this, a more or less homogeneous highly proficient Russian L1-German L2 group was tested on their use of several predictive cues, thus factors like German proficiency and AoA (late, after the age of seven years) were stable across experiments. All predictive cues tested also exist and are similarly realized in Russian, so the L2 speakers were familiar with these cues from their L1. Complementary offline tasks also tested their knowledge of German and only L2 speakers were included who showed sufficient offline knowledge, here knowledge about subject-verb agreement and case marking as well as lexical knowledge. Experiment 4 further controlled for a cognitive origin of L1/L2 differences. Here, a group of German-speaking children was included who had a lower memory capacity than the adult L2 group as assessed through an additional digit span task. A pre-test confirmed that the two learner groups were sensitive to the discourse-level cue investigated in Experiment 4. To answer the main research question, the four visual-world eye-tracking experiments conducted addressed the following questions:

1. Do L2 speakers of German use the lexical-semantics of verbs to predict the animacy of the upcoming direct object noun?
2. Do L1 and L2 speakers of German use verb number marking to predict the number feature of the upcoming subject?
3. Do L1 and L2 speakers of German use case marking on the first post-verbal argument to predict the thematic role of the final argument?
4. (a) Do child L1 and adult L2 speakers use implicit causality information in the same way as adult L1 speakers to predict upcoming reference?
   (b) Do the three groups differ in the onset of the (anticipatory) effect?

Under the assumption that L2 speakers' predictive abilities were selectively limited, a positive outcome was expected for (1) with no L1/L2 difference, but possible differences were expected for the following cues, which require either a morphosyntactic operation (number marking in (2) and case marking in (3)) or the integration of information sources from different linguistic domains (4). Below, I will discuss the experimental results as related to these aspects.

To recapitulate the prognosis based on the hypotheses described under section 3.2: According to the RAGE hypothesis, L1 and L2 speakers mainly differ in their ability to generate expectations. However, the authors already discussed the constraints of the hypothesis and as stated above, this claim has already been challenged by previous studies. According to L2 processing accounts assuming that L1 and L2 processing are basically the same, L2 speakers should not differ from L1 speakers in terms of processing mechanisms, but both might be affected by the same factors mentioned under subsection 3.2.2, for example the availability of cognitive resources and the accuracy/consistency of lexical information. Kaan (2014) proposed this specifically for prediction. Following the Lexical Bottleneck Hypothesis, differences between L1 and L2 speakers, for example in constructing predictive agreement relations, can be attributed to the bilingual mental lexicon. The SSH assumes that L1 and L2 speakers have the same processing mechanisms available but differ in the way they make use of information sources, especially grammatical information. They may also show difficulties in the real-time integration of cues. The IH assumes a persisting difficulty with the syntax-discourse interface, probably due to differences in resource allocation.

The four visual-world eye-tracking experiments presented were designed in such a way that it was possible to assess whether a source of information was used predictively by the groups tested. Here this was done by defining a critical time window, during which the probability that a particular AOI was fixated should significantly increase compared to another AOI and/or should differ between experimental conditions. The four experiments, as shown in Table 9.1, tested the following sources of information and levels of prediction:

	source of information	**specific level of prediction**	**overall level of prediction**
Exp. 1	lexical-semantics of verbs	direct object noun (animate or inanimate)	semantic feature of a noun
Exp. 2	verb number marking	subject (singular or plural)	grammatical feature of a noun
Exp. 3	case marked on 1st post-verbal argument	thematic role (Theme or Recipient)	syntax-semantics mapping
Exp. 4	argument and discourse structure	coreference (NP1 or NP2)	syntax-semantics-discourse mapping

Table 9.1: Experiment overview including the source of information and level of prediction

Experiment 1 and Experiment 2 tested the pre-activation of a particular feature of an upcoming noun based on information given at the verb. As expected, the results for Experiment 1 showed that, based on the lexical-semantics of the verb, the semantic feature animacy could be pre-activated by both L1 and L2 speakers. Hence, Experiment 1 established that prediction based on lexical-semantic information was possible for the L2 group. Nevertheless, the groups showed a different onset of the anticipatory effect with a later onset and less certainty in the L2 group.

Experiment 2 investigated the use of verb number marking to predict the number feature of an upcoming subject. Here, the expectations were less clear because subject-verb agreement, although a highly available cue, is not always reliable, unlike number agreement within NPs. In German like in English, the verb can precede or follow the subject and grammatical and real-world number can mismatch. An effect of condition indicating an anticipatory effect for the critical window, which had a similar length to that in Experiment 1, was absent. Only with the onset of the quantifier did the L1 group, but, at least initially not the L2 group, show a significant difference between the singular and plural conditions. However, the results had to be interpreted with caution. As discussed under section 6.4, it is unclear how they were affected by the experimental design and method. Hence, it remains a subject for future research whether verb number marking serves as a predictive cue in German, a morphologically richer language than English, where a predictive use was found in a group of children as young as three years but findings were mixed for adult native speakers.

Experiment 3 tested the prediction of the thematic role of the final argument based on case marking on the first post-verbal argument after a ditransitive verb. Both groups demonstrated anticipation of the target argument, the Theme, for the canonical structure ‘Dat > Acc.’ Over time, participants across groups were less likely to fixate the Theme for the non-canonical structure ‘Acc > Dat.’ The difference in the time course between the two word order conditions had its onset in the critical time window, during which the first post-verbal argument was heard, and, for the L1 group, resulted in an overall effect of condition in the post-critical time window. In the post-critical time window, an L1/L2 difference emerged: Whereas the L1 group quickly recovered from the initial competition between the accusative object/Theme and dative object/Recipient and correctly identified the final argument in the non-canonical condition, the competition persisted in the L2 group. Moreover, the likelihood that the Theme was fixated in the canonical condition further developed over time in the L2 group, indicating less certainty. Hence, the results from Experiment 3 rather showed an L1/L2 difference in reactive processing, indicating that the L2 speakers needed more time to integrate incoming information and were affected by the non-canonical word order to a greater extent than the L1 speakers. Importantly, both groups became aware that the target argument in the canonical condition was implausible for the non-canonical condition within the critical time window. A prior study in German (Hopp, 2015) found that the L2 group, unlike the L1 group, anticipated the Patient of the canonical structure irrespective of the case marked on the NP1. Hence, in contrast to this study, which tested the use of case marking via object topicalization, Experiment 3 showed that L2 speakers were sensitive to the case cue and used it predictively.

Experiment 4 tested the prediction of upcoming reference, i.e., prediction at the discourse-level. Anticipatory eye-movements towards the biased antecedent within the time window before the disambiguating information was given demonstrated that the child and adult L1 groups successfully used the IC information to establish coreference with either the NP1 or the NP2 referent of the matrix clause depending on the respective verb bias. In both L1 groups, there was a significant difference in the likelihood that the NP1 referent was fixated between verb bias conditions; the onset of divergence was only slightly delayed in the child group. For the L2 group, an anticipatory effect was confined to the NP1-biasing verbs. If their first preference to resolve the pronoun towards the subject/NP1 was supported by the IC information of the verb, the effect was ‘boosted.’ The onset of divergence was only towards the end of the critical window for an anticipatory effect, seen in the growth curve analysis as an effect of linear time on the likelihood that the NP1 referent was fixated for the NP1-biasing verbs. Although only evident in the separate analyses but not the group model, there is an indication that the L2 speakers allocated the weighting of information sources differently, being more likely to resolve an ambiguous pronoun with the subject/NP1, at least initially. Since the children were shown to have a lower (working) memory capacity than the adult L2 speakers, the findings are unlikely to be explained in terms of resource deficits, but more likely in terms of L1/L2 differences.

Consistent with assumptions of the RAGE hypothesis, Experiment 1 and Experiment 4 (to some extent also Experiment 2) showed at least subtle differences between L1 and L2 speakers. However, I argue that these differences further vary in their nature and can be separated into more individual and selective differences. In addition, the results from Experiment 3 indicated that these differences show up not only in predictive but also in reactive processing.

9.1.1 Evidence for individual differences

Results of previous experiments (section 4.3) and the current experiments demonstrate that L2 speakers' predictive processing abilities are not generally limited in such a way that prediction is impossible for L2 speakers. They can, for example, predict an upcoming noun and its semantic or syntactic features based on verb semantics or prior context. Hence, at least for certain information sources/levels of prediction, L1 and L2 prediction is similar. However, results also provide evidence for an influence of more general, individual factors: The later onset of an anticipatory effect in Experiment 1, which functioned as a control experiment in which prediction was expected to be observed in both groups, can be explained by factors influencing L2 processing as proposed by L2 processing accounts that explain L1/L2 differences in terms of weaker links, more diffuse lexical representations or slower lexical access. A different onset was also reported by Ito, Pickering, and Corley (2018), who attributed this to resource limitations in the L2 group. Another observation, also already reported in previous studies (e.g., Mitsugi, 2018; Peters et al., 2015), was that L2 speakers seemed less certain. Thus, a later onset may also be conflated with less certainty. The finding that L2 speakers benefited less from prediction than L1 speakers as shown in the post-critical window in the study by van Bergen and Flecken (2017) and Experiment 3 of the current thesis further supports the view that L2 processing might be characterized by less certainty. One may speculate that the reason for this originates in the bilingual mental lexicon (see subsection 3.2.2).

Another factor that potentially affected the results in previous studies was L1-L2 similarity. Although there might be multifarious reasons why Experiment 3, for the first time, found prediction based on case marking in an L2 group, it seems likely that familiarity with case marking played a role. Molinaro et al. (2017) found that even balanced early bilinguals demonstrated an influence of the properties of the first learned language. The authors note that "early language exposure biases the way in which different cues are weighted to pursue optimal prediction mechanisms in the new experience settings" (p. 72). They further emphasize, however, that L2 learners are able to use a cue predictively in their L2 after enough exposure (in line with Kaan, 2014), even if it is absent in the L1 (in line with Foucart et al., 2014), and prediction might be available to them as a learning mechanism, as has been proposed for children (see subsection 3.1.3). Here, all sources of information tested existed both in Russian and German. Follow-up experiments might include a further L2 group whose L1 does not realize case but relies on other cues like word order, or marks thematic roles differently.

9.1.2 Evidence for selective differences

We have seen that L2 speakers' ability to predict may in fact be reduced, but not because it is a processing mechanism that is not or is less available to L2 speakers, but because their processing is affected by a range of factors, some of them L2-specific (e.g., language similarity and experience). I argue that in addition to more general factors that can lead to individual differences, for example a predictive effect in the L2 speaker who is familiar with gender from his L1 but not for the L2 speaker who is not, there is also evidence for selective differences. The results of the experiments testing prediction based on a morphosyntactic cue, particularly the results for Experiment 3 on case marking, do not support that the use of grammatical cues per se is problematic. Note that based on the findings of previous studies (particularly Hopp, 2015; Kaan et al., 2016; Mitsugi & MacWhinney, 2016), less robust use of grammatical information was expected. This is also interesting in light of the L2 processing accounts discussed. According to the Lexical Bottleneck Hypothesis (Hopp, 2018), the mapping of lexical to syntactic information should form a 'bottleneck' for L2 speakers. Here, however, the L2 group was found to access the meaning of the first post-verbal argument, recognize the case marking, assign a thematic role and, on this basis, predict the upcoming thematic role. Only later did L1/L2 differences emerge. In line with Grüter et al. (2016), Grüter et al. (2018a) and prior assumptions regarding L2 processing (Clahsen & Felser, 2006a; Cunnings, 2017), I argue that the results show a different weighting of cues in L1 and L2 speakers. If information sources had to be integrated, L1/L2 differences emerged which were not restricted to the anticipatory process and cannot be explained in terms of a reduced ability to generate expectations, resource limitations and/or an inconsistent representation of lexical information.

When different information sources had to be integrated, as in the use of implicit causality, L2 speakers were found to rely on surface-level information like first-mention and/or subjecthood, at least early during processing. Typically, the first-mentioned referent and subject is also the Agent, a likely causer, and, hence, a likely referent of a subject pronoun. The findings support assumptions of both the SSH and the IH. Similarly, word order variation, another surface-level cue, had a more severe effect on the L2 group, a finding better accounted for by the SSH than the IH. Interestingly, the effect spilled over into the post-critical window and did not immediately lead to a difference in anticipation but rather affected L2 speakers' later sentence interpretation and re-analysis (see also the ERP difference between L1 and L2 speakers for the ungrammatical conditions in Kaan et al., 2016). Note that also the findings by

Hopp (2015) for the English L1-German L2 speakers can be interpreted as a problem with the integration of cues, here case marked on the NP1 together with the lexical-semantic information of the verb. The reason why the L2 speakers were found to rely on verb semantics and word order only might be, because these cues receive more weight in L2 processing, which is likely to be enhanced by a factor like language similarity, which probably biases cue weighting as stated by Molinaro et al. (2017). Unlike German, English relies on word order for thematic role assignment and does not mark case on lexical NPs. A difference in the relative weighting of cues has also been proposed by Grüter et al. (2016) for the use of event structure and by Grüter et al. (2018a) for the use of classifiers. Grüter et al. (2016) found that the L2, unlike the L1 group, made use of verb semantics but not grammatical aspect to predict upcoming reference. Grüter et al. (2018a) found that the L2 group relied more on the semantics of classifiers, whereas for the L1 group form class was the primary cue. Hence, if information from different linguistic domains needs to be integrated, L2 speakers are likely to put more weight on semantic and/or surface-level information, i.e., non-grammatical information, rather than grammatical information. The learner comparison in Experiment 4 further supports the assumption that the L1/L2 difference is not the result of resource limitations but rather resource allocation.

The results of Experiment 3 show that it is not only prediction that is affected by the different weighting of cues, but probably also L2 speakers' re-analysis (see also Jacob & Felser, 2016; Pozzan & Trueswell, 2016). This finding further indicates that differences in predictive abilities are not the dominant source of L1/L2 differences, although prediction as a fast-operating mechanism is likely to be affected. Recently, Cunnings (2017) argued for increased interference from an initially assigned interpretation in L2 speakers, who are susceptible to such interference during memory retrieval operations. It remains a subject for future research to what extent this assumption also applies to non-canonical structures. However, it might account for the findings in Experiment 4, which clearly involved a retrieval operation.

9.2 Critical notes & future directions

While developing the experiments, certain decisions were made, for example, in terms of the stimuli presentation. In the following, I will discuss those points I think are worth mentioning, either in light of the interpretation of the results or with respect to future research. I will further discuss possible ways to follow up on the results obtained.

In the current experimental designs one of the objects in the visual display functioned as a gaze neutralizer. Thus, instead of a fixation cross another AOI was used to attract the eye-movements away before the predictive cue was heard in order to minimize baseline effects. As it turned out in Experiment 1 and Experiment 2, the proportion of looks for the gaze neutralizer was still relatively high, hence leading to noise in the data. Another (potential) problem here was that the target at the same time was also the last word of the sentence. The uncertainty in the L2 group observed in Experiment 1 might be associated with final sentence interpretation, where, towards the end, all candidates visually displayed were again considered. Some evidence for this explanation comes from Experiment 4, where the proportion of looks for the target after the onset of the disambiguating information did not visibly differ between groups. In Experiment 4, all pictures shown were further mentioned during the sentence, unlike in the first three experiments. In Experiment 4, the natural pause after the matrix clause probably aided in neutralizing the participants' gaze here, as intended. Therefore, depending on the structure under investigation, the use of a certain experimental design appears more or less useful. One possibility to avoid that the participants' gaze is still directed to another AOI upon encountering the critical information, is to change the visual display and, perhaps, also when it is presented. In the VW experiments conducted by Rommers et al. (2013) and Ito, Pickering, and Corley (2018) the visual display was presented 500 ms (Rommers et al., 2013) or 1000 ms (Ito, Pickering, & Corley, 2018) before the target word onset and only showed one critical object in the respective experimental condition and three distractors. With the exception of Experiment 4, where the characters were introduced beforehand, it was not controlled whether participants would name the pictures accordingly. However, here the interest was not in word-specific prediction or, related to this, whether participants would use a word with a particular gender or phonological onset. For Experiment 1 and Experiment 2 the critical pictures clearly displayed the investigated feature, i.e., animate/inanimate and singular/plural. It was further controlled that the characters mentioned together in Experiment 3 differed in their profession, age and/or gender so as not to be confused.

In Experiment 2, a baseline effect complicated the data analysis and interpretation. This may show that eye-tracking is not the best method for comparing singular and plural. To follow up on the question of how reliable subject-verb agreement is as a cue, one may use the ERP method. Note, however, that then another problem arises, namely, how to tease apart prediction from rapid integration. For example, sentences like those that were interspersed to avoid a strategic use in Experiment 2 could be tested in an ERP experiment, see (104) below.

(104) In front of the window there is a ...

a. bird.

b. flight of birds.

c. *birds.

For the sentence version (104-b) one may expect to find a different ERP response starting at the offset of *birds* than for (104-a). However, there should be a difference from a continuation like (104-c), which is ungrammatical. A further contrast with a cue that is probably more reliable is needed to be able to draw conclusions about cue reliability like, for example, lexical gender in German.

This thesis could only comment on a selection of predictive cues. Future work may examine more linguistic sources of information (e.g., voice) and/or levels of prediction that may pose a challenge to language learners, also including other languages. The factor language similarity or cross-linguistic experience and its influence on prediction in L2 processing was mentioned several times. However, the nature of L1 transfer was not discussed in detail because it was beyond the scope of this thesis. Note though that it appears to be an interesting aspect and research so far lacks more L1-L2 comparisons in which a cue is not simply absent in the L1 of the L2 group but is realized differently. In the current experiments, the focus was on the comparison between adult L1 speakers of German who were raised in a monolingual environment during the first years of their life and adult L2 speakers who had learned German later in life and had reached a high proficiency level. In Experiment 4, German-speaking children were also tested, because no information on their online processing of IC information was available and because the comparison between both learner groups enabled me to control for possible effects of cognitive resource limitations.

For two of the cues investigated here, the use of the lexical-semantics of verbs and the prediction of the number feature of a noun based on verb number marking, data for children already existed. Previous experiments on German had also tested children's use of case marking, however only for SVO and OVS structures. It would be interesting to also test German-speaking children's use of case marking to predict an upcoming thematic role in double object structures as investigated in Experiment 3, as it might be a structure difficult for children up to the age of six years; see the results of an elicited imitation task by Drenhaus and Féry (2008) discussed under section 7.4. Moreover, a more thorough investigation of the interaction of different cues like case marking, word order and animacy seems promising. In the materials for Experiment 3, the animacy of the first post-verbal argument varied; a post-hoc analysis revealed that prototypical animacy features, see the example in (105), might facilitate the processing of non-canonical structures.

(105) Die Sekretärin gibt das piepende Kopiergerät schließlich dem Reparaturdienst.
The secretary gives [the beeping photocopier]-*Acc* finally [the repair service]-*Dat*
'The secretary finally gives the beeping photocopier to the repair service.'

As further discussed under section 7.4, it might be interesting to change the experimental design so it is closer to the one in a previous experiment conducted by Mitsugi and MacWhinney (2016), who detected no difference between the canonical and scrambled structures. Finally, a comparison between L2 groups with different L1-L2 typologies can shed light on the role language similarity plays in prediction. Since no L2 group without any proper case marking system in their L1 was tested in Experiment 3, the potential role of cross-linguistic experience remains speculative for the time being.

Chapter 10

Conclusion

The aim of the current thesis was to shed more light on the predictive processing abilities of late bilinguals. The research questions were motivated by the RAGE hypothesis put forward by Grüter et al. (2014). Since its proposal, this hypothesis has laid the foundation of a more thorough investigation of prediction in L2 speakers, as shown in the literature review, with most studies having been published within the last three years. Here the RAGE hypothesis served as a starting point to systematically test the interplay of prediction and the linguistic domain in L2 versus L1 speakers. Based on recent findings discussed under chapter 4 and my experimental results, I draw three main conclusions.

Prediction in L2 processing is possible. If we set aside subtle L1/L2 differences, we see that L2 speakers have been shown to predict upcoming linguistic information. For example, studies testing L2 speakers' use of the selectional restrictions of verbs robustly show that they use this information to fixate the picture showing the upcoming noun before they hear it. This finding could be replicated in Experiment 1 in the current thesis. Here, the level of prediction was more systematically controlled for than in previous studies by using verbs intended to restrict the selection of upcoming nouns to the semantic category animate or inanimate. Moreover, L2 speakers' prediction is not restricted to one linguistic domain, for example lexical-semantics. Experiment 3 demonstrated that highly proficient L2 speakers were even able to map the case marked on an argument to a thematic role in a timely enough manner to show a difference in the time course between two word order conditions before the onset of the final argument, indicating that they predicted the upcoming thematic role. Also in Experiment 4 on the predictive use of implicit causality information, L2 speakers were found to predict upcoming reference. However, the source(s) of information they used to generate their expectation seemed to be different from the sources of information used by the L1 groups, resulting in a different pattern. This leads me to the next conclusion.

Prediction in L2 processing is 'reduced.' As proposed by the authors' of the RAGE hypothesis, many experiments, including Experiment 1 and Experiment 4 in the current thesis, show major or at least subtle differences in the predictive processing abilities of L1 and L2 speakers. I argue that these differences have different origins: (i) Some of the L1/L2 differences observed can be explained in terms of more individual differences. These can include L2 proficiency and/or cross-linguistic differences as shown, for example, by van Bergen and Flecken (2017). A recurrent finding in visual-world eye-tracking is that L2 speakers appear to be slower/less certain than L1 speakers. Although further research is needed here, it seems likely that this is due to characteristics of the bilingual mental lexicon as suggested by Hopp (2018). Others have further argued for resource limitations in L2 processing that would affect their predictive abilities (e.g., Kaan, 2014). Ito, Pickering, and Corley (2018), for example, take the absence of phonological prediction in the L2 group as an indication of resource limitations. However, I argue that such individual differences cannot account for the whole range of findings. The learner comparison in Experiment 4 showed that cognitive resource limitations were unlikely to be the source of L1/L2 differences. The children's pattern largely resembled the adult L1 pattern, although a digit span task confirmed that they had a lower memory capacity than the L2 group. (ii) Hence, there was also evidence for selective differences. Previous findings (Grüter et al., 2016; Grüter et al., 2018a) and the findings for Experiment 4 support the assumption that L2 speakers show a different weighting of cues during real-time processing (Clahsen & Felser, 2006a; Cunnings, 2017). Note that L1/L2 differences here must reflect differences in processing and it is unlikely that they reflect L2 speakers' L2 knowledge as complementary offline tasks checked that they knew the predictive cues. In Experiment 4, the L2 group, unlike the L1 groups, showed an initial preference for predicting upcoming reference with the subject/NP1 of the matrix clause. Grüter et al. (2016) found that, unlike the L1 group, L2 speakers used verb semantics but not grammatical aspect to predict upcoming reference. To conclude, if information sources have to be integrated to form a prediction, L2 speakers rely more on semantic and surface-level information than L1 speakers.

L1/L2 differences are not restricted to predictive processing. Experiment 3 showed a predictive use of case marking across groups, however L1/L2 differences emerged in the post-critical window, which were indicative of differences in their reactive processing abilities. It appears that although prediction as a fast operating mechanism is likely to be affected in L2 processing, it is not the dominant source of L1/L2 processing differences. There is a greater indication that L2 speakers have difficulty with the real-time integration of cues. The L2 group in Experiment 3 was more affected by the non-canonical word order than the L1 group, again indicating that they relied more on a surface-level cue during real-time processing than the L1 group.

References

Abutalebi, J., & Clahsen, H. (2018). Critical periods for language acquisition: New insights with particular reference to bilingualism research. *Bilingualism: Language and Cognition*, *21*(5), 883–885.

Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, *46*, 219–227.

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1–L2 similarity and structural distance. *Second Language Research*, *30*(3), 275–306.

Altmann, G. T. M. (2002). Learning and development in neural networks – The importance of prior experience. *Cognition*, *85*(2), B43-B50.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*(1), 55–71.

Altmann, G. T. M., & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*(4), 583–609.

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, *28*(4), 525–560.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, *31*(2), 137–162.

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

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76*(1), B13–B26.

Atkinson, E., Wagers, M. W., Lidz, J., Phillips, C., & Omaki, A. (2018). Developing incrementality in filler-gap dependency processing. *Cognition*, *179*, 132–149.

Au, T. K.-F. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language*, *25*(1), 104–122.

Bader, M. (1994). Syntactic-function ambiguities. *Folia Linguistica*, *XXVIII*(1-2), 5–66.

Bamyacı, E., Häussler, J., & Kabak, B. (2014). The interaction of animacy and number agreement: An experimental investigation. *Lingua*, *148*, 254–277.

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289.

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. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

Barr, D. J., Gann, T. M., & Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica*, *137*(2), 201–207.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Becker, M. (2017). *The Acquisition of Syntactic Structure: Animacy and Thematic Alignment*. Cambridge: Cambridge University Press.

Berends, S. M., Brouwer, S. M., & Sprenger, S. A. (2016). Eye-tracking and the visual world paradigm. In M. S. Schmid et al. (Eds.), *Designing Research on Bilingual Development* (pp. 55–80). Cham: Springer.

Blake, B. J. (1997). *Case*. Cambridge: Cambridge University Press.

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In M. Carreiras & C. Clifton (Eds.), *The On-line Study of Sentence Comprehension* (pp. 51–76). New York, NY: Psychology Press.

Boland, J. E. (2005). Visual arguments. *Cognition*, *95*(3), 237–274.

Bornkessel, I., Zysset, S., Friederici, A. D., von Cramon, D. Y., & Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *NeuroImage*, *26*(1), 221–233.

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436.

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, *2*(1), 1–12.

Bott, O., & Solstad, T. (2014). From verbs to discourse: A novel account of implicit causality. In B. Hemforth, B. Mertins, & C. Fabricius-Hansen (Eds.), *Psycholinguistic Approaches to Meaning and Understanding across Languages* (pp. 213–251). Cham: Springer.

Boxell, O., & Felser, C. (2017). Sensitivity to parasitic gaps inside subject islands in native and non-native sentence processing. *Bilingualism: Language and Cognition*, *20*(3), 494–511.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS one*, *5*(5), e10773.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, *136*, 135–149.

Brouwer, S., Sprenger, S., & Unsworth, S. (2017). Processing grammatical gender in Dutch: Evidence from eye movements. *Journal of Experimental Child Psychology*, *159*, 50–65.

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, *14*(3), 237–273.

Bruns, T. (2007). *Einführung in die russische Sprachwissenschaft*. Tübingen: Narr.

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(4), 1029–1040.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, *47*(1), 30–49.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272.

Chang, F., Kidd, E., & Rowland, C. F. (2013). Prediction is a by-product of language learning [Comment on Pickering and Garrod, 2013]. *Behavioral and Brain Sciences*, *36*(4), 350–351.

Chauncey, K., Grainger, J., & Holcomb, P. J. (2008). Code-switching effects in bilingual word recognition: A masked priming study with event-related potentials. *Brain and Language*, *105*(3), 161–174.

Chen, L., Shu, H., Liu, Y., Zhao, J., & Ping, L. (2007). ERP signatures of subject-verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*(2), 161–174.

Cheng, W., & Almor, A. (2017). The effect of implicit causality and consequentiality on nonnative pronoun resolution. *Applied Psycholinguistics*, *38*(1), 1–26.

Cheng, W., & Almor, A. (2019). A Bayesian approach to establishing coreference in second language discourse: Evidence from implicit causality and consequentiality verbs. *Bilingualism: Language and Cognition*, *22*(3), 456–475.

Choi, Y., & Trueswell, J. C. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, *106*(1), 41–61.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, *22*(6), 991–1060.

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42.

Clahsen, H., & Felser, C. (2006b). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*, 107–126.

Clahsen, H., & Felser, C. (2006c). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–570.

Clahsen, H., & Felser, C. (2018). Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, *40*(3), 693–706.

Clahsen, H., & Muysken, P. (1986). The availability of universal grammar to adult and child learners – A study of the acquisition of German word order. *Second Language Research*, *2*(2), 93–119.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.

Comrie, B. (2006). *Language Universals and Linguistic Typology: Syntax and Morphology* (2nd ed.). Chicago: The University of Chicago Press.

Contemori, C., & Dussias, P. E. (2018). Prediction at the discourse level in L2 English speakers: An eye-tracking study. In A. B. Bertolini & M. J. Kaplan (Eds.), *Proceedings of the 42nd Annual Boston University Conference on Language Development* (pp. 159–171). Somerville, MA: Cascadilla Press.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84–107.

Corrigan, R. (1994). Children's causal attributions to states and events described by different classes of verbs. *Cognitive Development*, *9*, 235–256.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Cozijn, R., Commandeur, E., Vonk, W., & Noordman, L. G. (2011). The time course of the use of implicit causality information in the processing of pronouns: A visual world paradigm study. *Journal of Memory and Language*, *64*(4), 381–403.

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

Crinean, M., & Garnham, A. (2006). Implicit causality, implicit consequentiality and semantic roles. *Language and Cognitive Processes*, *21*(5), 636–648.

Cubberley, P. (2002). *Russian: A Linguistic Introduction*. Cambridge: Cambridge University Press.

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, *20*(4), 659–678.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychological Bulletin & Review*, *12*(3), 453–459.

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass*, *8*(12), 631–645.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121.

de Swart, P., Lamers, M., & Lestrade, S. (2008). Animacy, argument structure, and argument encoding. *Lingua*, *118*(2), 131–140.

Dienes, Z., Altmann, G. T. M., & Gao, S.-J. (1999). Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. *Cognitive Science*, *23*(1), 53–82.

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition*, *20*(5), 917–930.

Dink, J. W., & Ferguson, B. (2015). *eyetrackingR: An R library for eye-tracking data analysis.* Retrieved from `http://www.eyetrackingr.com`

Donnelly, S., & Verkuilen, J. (2017). Empirical logit analysis is not logistic regression. *Journal of Memory and Language*, *94*, 28–42.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–619.

Drenhaus, H., & Féry, C. (2008). Animacy and child grammar: An OT account. *Lingua*, *118*(2), 222–244.

Dröge, A., Fleischer, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2016). Neural mechanisms of sentence comprehension based on predictive processes and decision certainty: Electrophysiological evidence from non-canonical linearizations in a flexible word order language. *Brain Research*, *1633*, 149–166.

Dunabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816.

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*(02), 353–387.

Ellsiepen, E., & Bader, M. (2018). Constraints on argument linearization in German. *Glossa: A journal of general linguistics*, *3*(1), 6.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*(1), 71–99.

Featherstone, C. R., & Sturt, P. (2010). Because there was a cause for concern: An investigation into a word-specific prediction account of the implicit-causality effect. *Quarterly Journal of Experimental Psychology*, *63*(1), 3–15.

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.

Felser, C., & Clahsen, H. (2009). Grammatical processing of spoken language in child and adult language learners. *Journal of Psycholinguistic Research*, *38*(3), 305–319.

Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, *34*(1), 67–98.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15.

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*, *69*(3), 165–182.

Ferstl, E. C., Garnham, A., & Manouilidou, C. (2011). Implicit causality bias in English: A corpus of 300 verbs. *Behavior Research Methods*, *43*(1), 124–135.

Forster, K., & Forster, J. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods*, *35*(1), 116–124.

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*(5), 1461–1469.

Foucart, A., Ruiz-Tada, E., & Costa, A. (2015). How do you know I was about to say "book"? Anticipation processes affect speech processing and lexical recognition. *Language, Cognition and Neuroscience*, *30*(6), 768–780.

Foucart, A., Ruiz-Tada, E., & Costa, A. (2016). Anticipation processes in L2 speech comprehension: Evidence from ERPs and lexical recognition task. *Bilingualism: Language and Cognition*, *19*(1), 213–219.

Frederiksen, J. R. (1981). Understanding anaphora: Rules used by readers in assigning pronominal referents. *Discourse Processes*, *4*, 323–347.

Gabriele, A., Fiorentino, R., & Alemán Bañón, J. (2013). Examining second language development using event-related potentials. *Linguistic Approaches to Bilingualism*, *3*(2), 213–232.

Gabriele, A., Fiorentino, R., & Johnson, A. (2015). Attentional control and prediction in native and non-native speakers. *Linguistic Approaches to Bilingualism*, *5*(4), 470–475.

Gambi, C., Gorrie, F., Pickering, M. J., & Rabagliati, H. (2018). The development of linguistic prediction: Predictions of sound and meaning in 2-to-5 year olds. *Journal of Experimental Child Psychology*, *173*, 351–370.

Gambi, C., Pickering, M. J., & Rabagliati, H. (2016). Beyond associations: Sensitivity to structure in pre-schoolers' linguistic predictions. *Cognition*, *157*, 340–351.

Garnham, A. (2001). *Mental Models and the Interpretation of Anaphora*. Hove: Psychology Press.

Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, *5*(3), 459–464.

Garvey, C., Caramazza, A., & Yates, J. (1976). Factors influencing assignment of pronoun antecedents. *Cognition*, *3*(3), 227–243.

Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, *27*, 699–717.

Gil, D. (2013). Numeral classifiers. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from `http://wals.info/chapter/55`

Gillon Dowens, M., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish – Evidence from Event Related Potentials. *Neuropsychologia*, *49*(7), 1651–1659.

Gillon Dowens, M., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, *22*(8), 1870–1887.

Goethe-Institut. (2011). *German placement test.* Accessed online, with kind permission from the Goethe-Institut.

Goikoetxea, E., Pascual, G., & Acha, J. (2008). Normative study of the implicit causality of 100 interpersonal verbs in Spanish. *Behavior Research Methods*, *40*(3), 760–772.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814.

Greene, S. B., & McKoon, G. (1995). Telling something we can't know: Experimental approaches to verbs exhibiting implicit causality. *Psychological Science*, *6*(5), 262–270.

Grüter, T., Lau, E., & Ling, W. (2018a). L2 listeners rely on the semantics of classifiers to predict. In A. B. Bertolini & M. J. Kaplan (Eds.), *Proceedings of the 42nd Annual Boston University Conference on Language Development* (pp. 303–316). Somerville, MA: Cascadilla Press.

Grüter, T., Lau, E., & Ling, W. (2018b). *How classifiers facilitate processing in L2 Chinese.* Talk presented at the International Symposium on Bilingual Processing in Adults and Children (ISBPAC), TU Braunschweig, 24–25 May.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215.

Grüter, T., Rohde, H., & Schafer, A. J. (2014). The role of discourse-level expectations in non-native speakers' referential choices. In W. Orman & M. J. Valleau (Eds.), *Proceedings of the 38th Annual Boston University Conference on Language Development* (pp. 179–191). Somerville, MA: Cascadilla Press.

Grüter, T., Rohde, H., & Schafer, A. J. (2017). Coreference and discourse coherence in L2. *Linguistic Approaches to Bilingualism*, *7*(2), 199–229.

Grüter, T., Takeda, A., Rohde, H., & Schafer, A. J. (2016). *L2 listeners show anticipatory looks to upcoming discourse referents.* Poster presented at the 41st Annual Boston University Conference on Language Development, Boston University, 4–6 November.

Guerry, M., Gimenes, M., Caplan, D., & Rigalleau, F. (2006). How long to find the cause? An on-line investigation of implicit causality in sentence production. *Quarterly Journal of Experimental Psychology*, *59*, 1445–1465.

Haegeman, L. M. V. (1996). *Introduction to Government and Binding Theory* (2. ed., repr ed.). Oxford: Blackwell.

Haider, H. (2017). Mittelfeld Phenomena (Scrambling in Germanic). In M. Everaert & H. C. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax. Vol. 3* (pp. 1–73). Oxford: Blackwell.

Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on language technologies (NAACL 01)* (pp. 1–8).

Härtl, H. (2001). *Cause und Change: Thematische Relationen und Ereignisstrukturen in Konzeptual-*

isierung und Grammatikalisierung (Vol. 50). Berlin: Akademie Verlag.

Hartshorne, J. K. (2014). What is implicit causality? *Language, Cognition and Neuroscience*, *29*(7), 804–824.

Hartshorne, J. K., O'Donnell, T. J., & Tenenbaum, J. B. (2015). The causes and consequences explicit in verbs. *Language, Cognition and Neuroscience*, *30*(6), 716–734.

Hartshorne, J. K., Pogue, A., & Snedeker, J. (2015). Love is hard to understand: The relationship between transitivity and caused events in the acquisition of emotion verbs. *Journal of Child Language*, *42*(3), 467–504.

Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, *28*(10), 1474–1508.

Hartshorne, J. K., Sudo, Y., & Uruwashi, M. (2013). Are implicit causality pronoun resolution biases consistent across languages and cultures? *Experimental Psychology*, *60*(3), 179–196.

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277.

Haspelmath, M. (2013). Ditransitive constructions: The verb 'give'. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from `http://wals.info/chapter/105`

Henry, N., Hopp, H., & Jackson, C. N. (2017). Cue additivity and adaptivity in predictive processing. *Language, Cognition and Neuroscience*, *32*(10), 1229–1249.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradim: a window onto emerging language comprehension. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for Assessing Children's Syntax* (pp. 105–124). Cambridge, MA: MIT Press.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, *22*(3), 369–397.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901–931.

Hopp, H. (2012). The on-line integration of inflection in L2 processing: Predictive processing of German gender. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th Annual Boston University Conference on Language Development* (Vol. 1, pp. 226–245). Somerville, MA: Cascadilla Press.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56.

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching*, *53*(3), 277–306.

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307.

Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second Language*, *17*, 5–27.

Hopp, H., & Lemmerth, N. (2017). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*(1), 171–199.

Hosemann, J., Herrmann, A., Steinbach, M., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Lexical prediction via forward models: N400 evidence from German Sign Language. *Neuropsychologia*, *51*(11), 2224–2237.

Huang, Y. T., Zheng, X., Meng, X., & Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language*, *69*(4), 589–606.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23-B32.

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93.

Huettig, F., & Mani, N. (2015). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. R., & Trueswell, J. C. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research*, *29*(6), 597–626.

Ito, A., Corley, M., & Pickering, M. J. (2018). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition*, *21*(2), 251–264.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*, 157–171.

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). On predicting form and meaning in a second language. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(4), 635–652.

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11.

Jackson, C. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning*, *58*(4), 875–909.

Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology*, 907–925.

Jacobs, J. (1988). Probleme der freien Wortstellung im Deutschen. *Sprache und Pragmatik*, *5*, 8–37.

Järvikivi, J., van Gompel, R. P. G., & Hyönä, J. (2017). The interplay of implicit causality, structural heuristics, and anaphor type in ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, *46*(3), 525–550.

Järvikivi, J., van Gompel, R. P. G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science*, *16*(4), 260–264.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*(4), 603–634.

Johnson, A. (2015). *Individual differences in predictive processing: Evidence from subject filled-gap effects in native an nonnative speakers of English* (Unpublished doctoral dissertation). University of Kansas.

Johnson, A., Fiorentino, R., & Gabriele, A. (2016). Syntactic constraints and individual differences in native and non-native processing of wh-movement. *Frontiers in Psychology*, *7*, 549.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*(2), 257–282.

Kaan, E., Dallas, A., & Wijnen, F. (2010). Syntactic predictions in second-language sentence processing. In C. J.-W. Zwart & M. d. Vries (Eds.), *Structure Preserved* (pp. 207–214). Amsterdam: John Benjamins Publishing Company.

Kaan, E., Kirkham, J., & Wijnen, F. (2016). Prediction and integration in native and second-language processing of elliptical structures. *Bilingualism: Language and Cognition*, *19*(1), 1–18.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113–147.

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

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2*(4), 647–670.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, *25*(1), 1–44.

Kempe, V. (1999). Processing of morphological and semantic cues in Russian and German. *Language and Cognitive Processes*, *14*(2), 129–171.

Kholodova, A., & Allen, S. (in press). The ditransitive alternation in German: Structural preferences and verb bias effects. In *Ditransitive constructions in Germanic languages.* Amsterdam: John Benjamins Publishing Company.

Kidd, E., Stewart, A. J., & Serratrice, L. (2011). Children do not overcome lexical biases where adults do: The role of the referential scene in garden-path recovery. *Journal of Child Language*, *38*(1), 222–234.

Kim, H., & Grüter, T. (2018). *Processing of remention biases in Korean learners of English.* Talk presented at the 43rd Annual Boston University Conference on Language Development, Boston University, 2–4 November.

Kim, H., & Grüter, T. (2019). Cross-linguistic activation of implicit causality biases in Korean learners of English. *Bilingualism: Language and Cognition*, *22*(3), 441–455.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127.

Koornneef, A. W., & Sanders, T. J. M. (2013). Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, *28*(8), 1169–1206.

Koornneef, A. W., & van Berkum, J. J. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, *54*(4), 445–465.

Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular-plural distinction. *Language Learning and Development*, *2*(1), 1–25.

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*(1), 23–42.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59.

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, *23*(5), 1230–1246.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the Brain* (pp. 190–207). New York: Oxford University Press.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470.

Kutas, M., van Petten, C., & Kluender, R. (2006). Psycholinguistics Electrified II (1994-2005). In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (pp. 659–674). Amsterdam and Boston: Elsevier/Academic Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Lago, S., & Felser, C. (2018). Agreement attraction in native and nonnative speakers of German. *Applied Psycholinguistics*, *39*(3), 619–647.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484–502.

Leiner, D. J. (2014). *SoSci survey.* Retrieved from `https://www.soscisurvey.de`

Lemmerth, N., & Hopp, H. (2017). Gender processing in simultaneous and successive bilingual children: Cross-linguistic lexical and syntactic influences. *Language Acquisition*, *26*(1), 21–45.

Lenerz, J. (1977). *Zur Abfolge nominaler Satzglieder im Deutschen* (Vol. 5). Tübingen: Narr.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Lew-Williams, C. (2017). Specific referential contexts shape efficiency in second language processing: Three eye-tracking experiments with 6- and 10-year-old children in Spanish immersion schools. *Annual Review of Applied Linguistics*, *37*, 128–147.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464.

Lim, J. H., & Christianson, K. (2013). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition*, *16*(3), 518–537.

Long, D. L., & de Ley, L. (2000). Implicit causality and discourse focus: The interaction of text and reader characteristics in pronoun resolution. *Journal of Memory and Language*, *42*(4), 545–570.

Lukyanenko, C., & Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition*, *146*, 349–370.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (2001). The competition model: the input, the context, and the brain. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 69–90). New York, NY: 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*(2), 127–150.

Mak, W., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language*, *54*(4), 466–490.

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake – But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847.

Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *Journal of Experimental Child Psychology*, *126*, 264–279.

Mannetti, L., & de Grada, E. (1991). Interpersonal verbs: Implicit causality of action verbs and contextual factors. *European Journal of Social Psychology*, *21*, 429–443.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, *57*(4), i-178.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is production: The missing link between language production and comprehension. *Scientific Reports*, *8*, 1079.

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*(4), 574–588.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372–380.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381–401.

McDonald, J. L., & MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit causality and gender. *Journal of Memory and Language*, *34*, 543–566.

McKoon, G., Green, S. B., & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *19*(5), 1040–1052.

Meisel, J. M. (2013). *First and Second Language Acquisition: Parallels and Differences* (Reprint ed.). Cambridge: Cambridge University Press.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494.

Mitsugi, S. (2018). Generating predictions based on semantic categories in a second language: A case of numeral classifiers in Japanese. *International Review of Applied Linguistics in Language Teaching*. doi: https://doi.org/10.1515/iral-2017-0118

Mitsugi, S., & MacWhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition*, *19*(1), 19–35.

Molinaro, N., Giannelli, F., Caffarra, S., & Martin, C. (2017). Hierarchical levels of representation in language prediction: The influence of first language acquisition in highly proficient bilinguals. *Cognition*, *164*, 61–73.

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (words): An electrophysiological study of code switching. *Brain and Language*, *80*(2), 188–207.

Müller, G. (1999). Optimality, markedness, and word order in German. *Linguistics*, *37*(5), 777—818.

Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, *86*(4), 314–329.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*(1), 11–28.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, *7*, e33468.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(3), 263–281.

O'Grady, W. (2008). The emergentist program. *Lingua*, *118*(4), 447–464.

Omaki, A., Davidson White, I., Goro, T., Lidz, J., & Phillips, C. (2014). No fear of commitment: Children's incremental interpretation in English and Japanese wh-questions. *Language Learning and Development*, *10*(3), 206–233.

Omaki, A., & Lidz, J. (2015). Linking parser development to acquisition of syntactic knowledge. *Language Acquisition*, *22*(2), 158–192.

Otten, M., Nieuwland, M. S., & van Berkum, J. J. A. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, *8*, 89.

Otten, M., & van Berkum, J. J. A. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, *45*(6), 464–496.

Özge, D., Kornfilt, J., Münster, K., Knoeferle, P., Küntay, A., & Snedeker, J. (2016). Predictive use of case markers in German children. In J. Scott & D. Waughtal (Eds.), *Proceedings of the 40th Annual Boston University Conference on Language Development* (pp. 291–303). Somerville, MA: Cascadilla Press.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*(3), 427–456.

Peters, R. E., Grüter, T., & Borovsky, A. (2015). Anticipatory and locally coherent lexical activation varies as a function of language proficiency. In D. C. Noelle et al. (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 1865–1870). Austin, TX: Cognitive Science Society.

Phillips, C., & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Linguistic Approaches to Bilingualism*, *5*(4), 409–453.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–110.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347.

Pickering, M. J., & Majid, A. (2007). What are implicit causality and consequentiality? *Language and Cognitive Processes*, *22*(5), 780–788.

Pozzan, L., & Trueswell, J. C. (2016). Second language processing and revision of garden-path sentences: A visual word study. *Bilingualism: Language and Cognition*, *19*(3), 636–643.

Primus, B. (1999). *Cases and Thematic Roles: Ergative, Accusative and Active*. Tübingen: Niemeyer.

Primus, B. (2012). *Semantische Rollen* (Vol. 12). Heidelberg: Universitätsverlag Winter.

Puga, J. L., Krzywinski, M., & Altman, N. (2015). Bayes' theorem. *Nature Methods*, *12*(4), 277–278.

Pyykkönen, P., & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, *57*(1), 5–16.

Pyykkönen-Klauck, P., & Crocker, M. W. (2016). Attention and eye movement metrics in visual world eye tracking. In P. Knoeferle, P. Pyykkönen-Klauck, & M. W. Crocker (Eds.), *Visually Situated Language Comprehension* (pp. 67–82). Amsterdam and Philadelphia: John Benjamins Publishing Company.

R Core Team. (2017). *R: A language and environment for statistical computing.* Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rabagliati, H., Gambi, C., & Pickering, M. J. (2016). Learning to predict or predicting to learn? *Language, Cognition and Neuroscience*, *31*(1), 94–105.

Riordan, B., Dye, M., & Jones, M. N. (2015). Grammatical number processing and anticipatory eye movements are not tightly coordinated in English spoken language comprehension. *Frontiers in Psychology*, *6:590*.

Robertson, E. K., Shi, R., & Melancon, A. (2012). Toddlers use the number feature in determiners during online noun comprehension. *Child Development*, *83*(6), 2007–2018.

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, *51*(3), 437–447.

Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language*, *38*, 150–176.

Rudolph, U. (2008). Covariation, causality, and language: Developing a causal structure of the social world. *Social Psychology*, *39*(3), 174–181.

Rudolph, U., & Försterling, F. (1997). Zur impliziten Kausalität in Sprache: Kriterien zur Selektion von Stimulusmaterial in Studien zur Verbkausalität: [The implicit causality in language: On the selection of stimulus materials in studies on verb causality]. *Zeitschrift für Experimentelle Psychologie*, *2*, 293–304.

Sauppe, S. (2016). Verbal semantics drives early anticipatory eye movements during the comprehension of verb-initial sentences. *Frontiers in Psychology*, *7:95*.

Saville-Troike, M., & Barto, K. (2012). *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.

Schad, D. J., Hohenstein, S., Vasishth, S., & Kliegl, R. (July 2018). *How to capitalize on a priori contrasts in linear (mixed) models: A tutorial.* Retrieved from `https://arxiv.org/abs/1807.10451v1`

Schipke, C. S., Knoll, L. J., Friederici, A. D., & Oberecker, R. (2012). Preschool children's interpretation of object-initial sentences: Neural correlates of their behavioral performance. *Developmental Science*, *15*(6), 762–774.

Schlesewsky, M., Bornkessel, I., & Frisch, S. (2003). The neurophysiological basis of word order variations in German. *Brain and Language*, *86*(1), 116–128.

Schumacher, P. B., Roberts, L., & Järvikivi, J. (2017). Agentivity drives real-time pronoun resolution: Evidence from German er and der. *Lingua*, *185*, 25–41.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.

Sekerina, I. A. (1997). *The syntax and processing of Russian scrambled constructions.* (Unpublished doctoral dissertation). City University of New York.

Sekerina, I. A. (2003). Scrambling and processing: Dependencies, complexity, and constraints. In S. Karimi (Ed.), *Word Order and Scrambling.* Malden, MA: Blackwell.

Sekerina, I. A., Stromswold, K., & Hestvik, A. (2004). How do adults and children process referentially ambiguous pronouns? *Journal of Child Language*, *31*(1), 123–152.

SensoMotoric Instruments GmbH. (February 2012). iView X System^{TM} Manual. Version 2.8 [Computer software manual].

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, *13*(3), 272–281.

Slabakova, R. (2008). *Meaning in the Second Language* (Vol. 34). Berlin: Mouton de Gruyter.

Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, *23*(3), 197–229.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299.

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism*, *1*(1), 1–33.

Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, *13*(2), 195–210.

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

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language*, *42*(3), 423–443.

Szewczyk, J. M., & Schriefers, H. (2011). Is animacy special? ERP correlates of semantic violations and animacy violations in sentence processing. *Brain Research*, *1368*, 208–221.

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, *68*(4), 297–314.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language information. *Science*, *286*, 1632–1634.

Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye movements and spoken language comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (pp. 863–900). Amsterdam and Boston: Elsevier/Academic Press.

Tewes, U. (1991). *Hamburg-Wechsler Intelligenztest für Erwachsene, Revision 1991: Handbuch und Testanweisung* (1st ed.). Bern: Huber.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, *27*(2), 173–204.

Traxler, M. J. (2002). Plausibility and subcategorization preference in children's processing of temporarily ambiguous sentences: Evidence from self-paced reading. *The Quarterly Journal of Experimental Psychology*, *55*(1), 75–96.

Trueswell, J. C., Sekerina, I. A., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition*, *73*, 89–134.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*(3), 285–318.

Tsang, C., & Chambers, C. G. (2011). Appearances aren't everything: Shape classifiers and referential processing in Cantonese. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(5), 1065–1080.

Ullman, M. T. (2001a). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(1), 105–122.

Ullman, M. T. (2001b). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews*, 717–727.

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and Context in Adult Second Language Acquisition* (pp. 141–178). Washington DC: Georgetown University Press.

Uszkoreit, H. (1986). Constraints on order. *Linguistics*, *24*, 883–906.

van Bergen, G., & Flecken, M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. *Journal of Memory and Language*, *92*, 26–42.

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

van Heugten, M., & Shi, R. (2009). French-learning toddlers use gender information on determiners during word recognition. *Developmental Science*, *12*(3), 419–425.

van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190.

van Valin, R. D., & LaPolla, R. (1997). *Syntax. Structure, Meaning and Function*. Cambridge: Cambridge University Press.

Vorster, J. (1984). Implicit causality in language: Evidence from Afrikaans. *South African Journal of Psychology*, *15*(2), 62–67.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237.

Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology*, *99*(2), 75–95.

Wicha, N. Y. 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), 1272–1288.

Wu, Y. J., & Thierry, G. (2010). Chinese-English bilinguals reading English hear Chinese. *The Journal of Neuroscience*, *30*(22), 7646–7651.

Yamamoto, M. (2006). *Agency and Impersonality: Their Linguistic and Cultural Manifestations* (Vol. 78). Amsterdam and Philadelphia: John Benjamins Publishing Company.

Yamashita, H. (1997). The effects of word-order and case marking information on the processing of Japanese. *Journal of Psycholinguistic Research*, *26*(2), 163–188.

Yang, C. D. (2016). *The Price of Linguistic Productivity: How Children Learn to Break the Rules of Language*. Cambridge, MA: MIT Press.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*(1), 1–14.

Zhang, L., & Knoeferle, P. (2012). Visual context effects on thematic role assignment in children versus adults: Evidence from eye tracking in German. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Society* (pp. 2593–2598). Austin, TX: Cognitive Science Society.

Appendix A

Experimental design, set-up and data preparation

A.1 The experimental design and set-up

In the following, I will describe the experimental design and set-up for the VW experiments. In Experiments 1, 2 and 4, the visual display contained three AOIs: a target, a competitor and a gaze neutralizer. The gaze neutralizer was used instead of a fixation cross to preserve the more natural processing in VW experiments. For the example (106-a) from Experiment 1, the gaze neutralizer was the picture of the subject or Agent, here a woman displayed below in the center, the target the picture of a cat, the only animate object not yet mentioned, and the competitor the picture of a blouse, an inanimate object. Experimental conditions were equally distributed across two lists using a Latin square design, so an item was seen only once by each participant. In Experiment 1, the display was the same for the sentence in (106-b), but here the picture of the blouse became the target.

(106) a. Die Frau füttert **die schwarze** Katze.
The woman feeds the black cat.

b. Die Frau bügelt **die schwarze** Bluse.
The woman irons the black blouse.

The positions of the target and competitor were switched across items/conditions since the top left position might be favored by participants (especially if their spoken languages are written from left to right), see Figure A.1. Thus, for another item with an animate target, the target would be displayed in the top right position. Critically, a participant's gaze was expected to shift from the bottom picture to one of the pictures displayed at the same distance above.

In Experiment 1, the critical question was whether the lexical-semantics of verbs is used to anticipate either an animate or an inanimate object. Highlighted in bold in the examples in (106) is the anticipatory time window. The noun has not yet been encountered, however based on prior information it can be predicted. The adjective extends the critical region. To avoid the length of the critical window being different between the conditions for an item, giving participants more time to anticipate in one of them, or there being an effect of prosody, the critical region in all experiments was identical. Further below, I describe how this was done by splicing this part of the sentence.

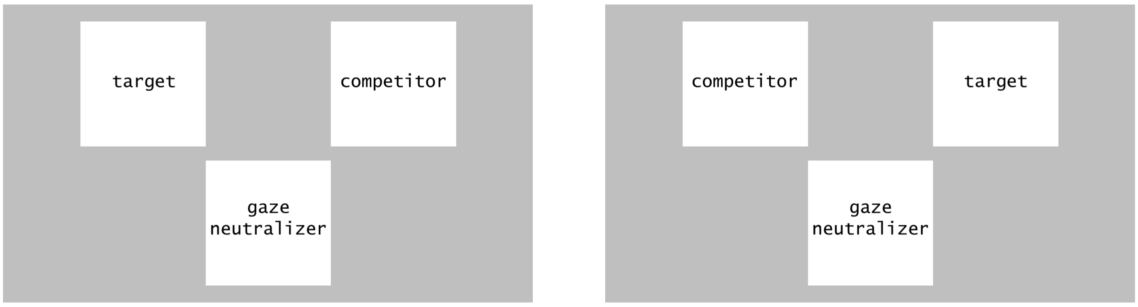


Figure A.1: Experimental design in this thesis: Areas of interest (AOIs)

For the analyses, the AOIs were coded in a principled way. Note that there are different options. For the growth curve analysis for Experiment 1 presented in subsection 5.3.2, looks to the two critical AOIs were coded as looks to the target versus looks to the competitor. In principle, one could also analyze looks towards one semantic category, for example the animate objects, depending on the condition. To demonstrate that the outcome is the same but has to be interpreted differently, I carried out the same analysis with the Elog-transformed fixation proportions for the animate AOI as dependent variable. In Table A.1 below the effect at the intercept seen in Table 5.4 in chapter 5, reflecting an advantage of the target over the competitor across conditions, now turns into an effect of condition: Participants across groups were less likely to fixate the animate AOI if they heard the sentences in the inanimate condition. Note that the *t*- and *p*-values (shown in bold) are the same as for the target-versus-competitor analysis.

	Estimate	Std. Error	*t*-value	*p*-value
Intercept (condition = animate)	0.5706	0.1867	3.057	0.0034 **
Condition (inanimate)	-0.9902	0.2693	**-3.676**	**0.0005** ***
Group	0.2543	0.3063	0.830	0.4096
Linear Time	-0.0932	0.2997	-0.311	0.7572
Condition x Group	-0.418	0.4564	-0.917	0.3627
Condition x Linear Time	-1.4878	0.4959	-3.000	0.0042 **
Group x Linear Time	0.0996	0.598	0.166	0.8684
Condition x Group x Linear Time	0.1395	0.9903	0.141	0.8886

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Formula: Elog ~ condition*group*(ot1) + (1 + condition*group| item) + (1 + condition*ot1 | subject)

Table A.1: Results of the growth curve analysis for the critical window in Experiment 1 (alternative)

Further note that only fixations on the two critical AOIs, for Experiment 1 the animate and inanimate objects, were included, because the interest lay in the relative difference. I decided to analyze the fixations on the singular AOI in Experiment 2 and the fixations on the NP1 referent in Experiment 4, so like above an effect of the experimental manipulation should be visible in terms of an effect of condition. For the analyses in Experiment 2 this was done to better interpret the data in light of the overall preference for the plural AOI. In Experiment 4 this was done because subject pronouns typically refer to the NP1, so the likelihood of fixations on the NP1 referent in the NP1-biasing contexts was taken as the baseline; in Experiment 3 the same was done with the canonical condition.

Experiment 3 differed from the experimental design described above. Here, four AOIs were presented. Again, one picture functioned as a competitor for another. Depicted for the example below were pictures of a man, a baby, a pacifier and a woman. During the NP *baby* participants were expected to move their eyes towards the target AOI, the picture of the pacifier in (107-a), but the picture of a plausible Recipient in (107-b).

(107)	a.	Der Vater	überreicht	dem schreienden	**Baby**	**vorsichtig**	den	Schnuller.
		[the father]-*Nom*	hands over	[the crying	baby]-*Dat*	carefully	[the	pacifier]-*Acc*
		‘The father carefully hands the pacifier to the baby.’						
	b.	Der Vater	überreicht	das schreiende	**Baby**	**vorsichtig**	der	Mutter.
		[the father]-*Nom*	hands over	[the crying	baby]-*Acc*	carefully	[the	mother]-*Dat*
		‘The father carefully hands the baby carefully to the mother.’						

Since the structure in (107-a) was the canonical one, it appeared useful to treat this experimental condition as the baseline and to analyze the probability of looks towards the Theme of condition (a), that is the same picture for an item, depending on experimental condition. As for the other experiments, only the two critical AOIs were considered. The exact procedures are described in the data preparation and analyses sections of the individual experiments.

In all experiments, participants had a preview time of 1000 ms, so they had enough time to scan the visual display before sentence onset. To take into account the time to initiate a gaze shift, 200 ms were added to the critical on- and offsets. Target fixations before or within the first 200 ms of the critical window for an anticipatory effect were not removed, but the trajectory of the effect was analyzed. The underlying assumption was as follows: If participants were already fixating the target picture, their attention should remain on this picture, whereas if they were fixating the gaze neutralizer or the competitor, they should shift their attention to the target (see also Barr et al., 2011).

A.2 Data preparation

Before setting up the experiment, the pictures and audio files were edited. In Experiments 1–3 non-colored line drawings were used (saved as .png), and in Experiment 4 photographs (saved as .jpg). The sentences for the experiments were recorded in a soundproof booth at the University of Potsdam. The 1000 ms preview time and the additional 800 ms after each sentence were included in the audio files as silence. In a further step, the critical time window for an anticipatory effect for each experimental item was cross-spliced, using Praat (Boersma, 2001). For the example in (106), see also Figure A.2, the article-adjective segment was copied from the sentence in the animate condition, then the respective segment was removed from the sentence in the inanimate condition and replaced by that from the animate condition. This was done for all items/conditions including the identical segment. From which sentence condition the spliced segment was taken depended on the acoustical fit. Crucially, no participant reported that the audio files sounded weird in any way. I also used Praat to determine the on- and offsets.

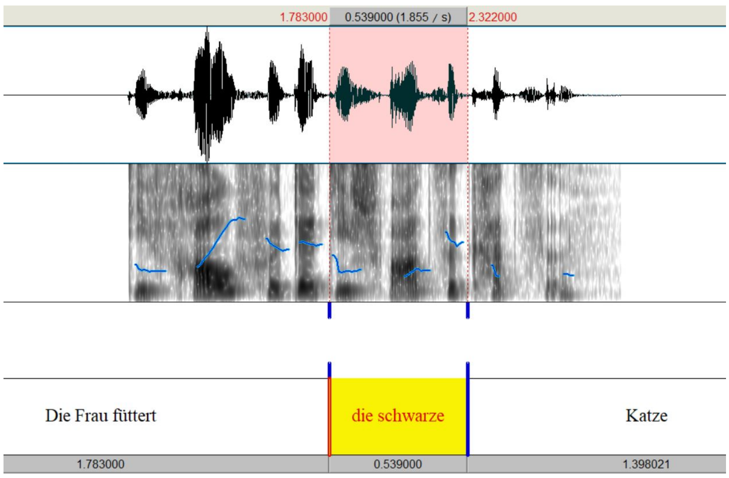


Figure A.2: Audio editing in Praat

The testing took place in a room without natural sunlight, thus avoiding changes in illumination. For all the experiments, a remote eye-tracker from SMI was used that had a sampling rate of 120 Hz, so the camera took a sample approximately every 8.33 ms. Both eyes were tracked and their movements averaged (input filter setting in iViewX). No chin rest was used, but participants were instructed to avoid body movements. They were informed about breaks in between. Each test session started with a calibration, that is, a reference measurement, enabling the eye-tracking system to relate the eye position to points in the calibration area that corresponds to the monitor screen. A validation then checked the accuracy of the eye-tracking system (SensoMotoric Instruments GmbH, February 2012).

After data collection, the data were exported from the proportions of looks module in the SMI data analysis software program BeGaze. Since the interest lay in the fixation proportions, only fixations were exported (alternatively, one can also choose the raw data export, including saccades and blinks). The output included columns as shown in Table A.2. Here, the first two observations or samples of the 20th trial for participant L1_02, the second participant in the L1 group, are shown. This participant received the second list, so she saw the stimulus in the inanimate (ia) condition. At the start of the trial she fixated none of the specified AOIs but rather the background of the screen as indicated by "-". Further columns contain the exact x- and y-coordinates.

Subject	Task	Trial	Stimulus	audio playback	Gaze X [px]	Gaze Y [px]	Time [ms]	Type	AOI
L1_02	list2	Trial020	01ia_lexical	01_ia_spliced.wav	494.2	551.4	6	Fixation	-
L1_02	list2	Trial020	01ia_lexical	01_ia_spliced.wav	503.1	557.9	14	Fixation	-

Table A.2: Example of a BeGaze output file

In a next step, the data were merged with the on- and offset information that had previously been determined in Praat, since the interesting question was which AOI was fixated within a critical region. For the experiment taken as an example here, Experiment 1, this would be the anticipation of the post-verbal argument or direct object noun, so the critical region lasted from the onset of the article up to the offset of the adjective preceding the noun. To take into account eye-movement latency, 200 ms were added.

det_on_200	adj_off_200
2146	2685

Table A.3: Adding the time information for the eye-tracking analysis

For further data preparation and the statistical analyses, the eyetrackingR package developed by Dink and Ferguson (2015) was used. This package provides all tools necessary to analyze visual-world data. However, to use it, the data have to be in a particular format, so separate columns for each AOI and trackloss were defined, using the information from the AOI column in the BeGaze output file. Since in the first sample the participant fixated the background of the screen, this fixation was counted as trackloss. Actual trackloss, e.g., due to blinks, is just missing samples and was not included in the output. Here, the decision was to also treat other non-AOI looks as missing.

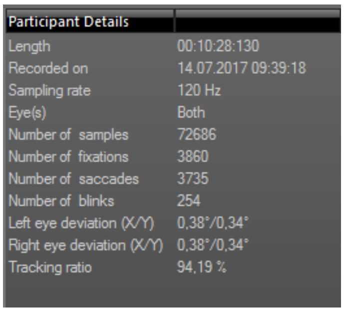
agent	animate	inanimate	trackloss
FALSE	FALSE	FALSE	TRUE

Table A.4: Adding columns for each area of interest for the eye-tracking analysis

The package further provides options for examining, for example, the trackloss per trial and participant. For the current experiments, participants were excluded based on the overall tracking ratio indicated in BeGaze, if below 80% overall, or on the calibration/validation measure (see Figure A.3) and the experimenter's observations during testing.[46] Single trials were excluded when a participant started

[46]One participant, for example, was observed to only look at the picture displayed in the top left position of the screen

coughing etc. Then this trial was excluded based on the experimenter's notes. Further data cleaning procedures are explained in the data preparation section for the individual experiments.



Participant Details	
Length	00:10:28:130
Recorded on	14.07.2017 09:39:18
Sampling rate	120 Hz
Eye(s)	Both
Number of samples	72686
Number of fixations	3860
Number of saccades	3735
Number of blinks	254
Left eye deviation (X/Y)	0,38°/0,34°
Right eye deviation (X/Y)	0,38°/0,34°
Tracking ratio	94,19 %

Figure A.3: Participant details in BeGaze

To specify a critical region, the eyetrackingR package comes with the subset_by_window function. The on- and offset information per stimulus was added before, so these columns could now be referred to as the trial start and end times. The onset of the critical region was aligned to zero and, for the statistical analyses, observations before and after were removed (for the time course graphs it was useful to keep these data). Next, the time course was defined with the help of the make_time_sequence_data function. Here, one or more predictor columns can be specified, e.g., the experimental condition and language group, as well as the AOI the time course will be defined for. At this step, the data points are binned, for the current experiments time bins of 50 ms were used.

```
timecourse_lexical <- make_time_sequence_data(critical_window_clean_2AOIs, time_bin_size = 50,
                                              predictor_columns = c("Group","cond"),
                                              aois = c("target")
)
```

Figure A.4: Defining the time course in R

For the growth curve analysis, trials were left distinct and were not aggregated as this is preferable for mixed-effects models calculated with the lmer function in R (R Core Team, 2017) as noted in the description for the make_time_sequence_data function. The data frame generated through this function includes columns with the proportion of looks and other transformations, such as the empirical logit (Elog) transformation that was used for the statistical analyses here, and orthogonal time polynomials for the calculation of a growth curve analysis.

throughout the experiment. Since it was intended that participants' gaze patterns should not be influenced in any way, this behavior was not corrected but this participant was excluded from the data set due to not having followed the instructions. When asked later, s/he said s/he was able to see the other pictures out of the corner of her/his eyes.

Appendix B

Experiment 1–3: Materials and additional information

B.1 Experiment 1: Experimental items

No	**subject/Agent verb**	CRITICAL WINDOW	**direct object noun**	cloze probability
1a	Die Frau füttert The woman feeds	die schwarze the black	Katze. cat.	0.85
1b	Die Frau bügelt The woman irons	die schwarze the black	Bluse blouse.	0.37
2a	Die Mutter badet The mother bathes	das dreckige the dirty	Kind. child.	0.83
2b	Die Mutter spült The mother cleans	das dreckige the dirty	Geschirr. dishes.	0.89
3a	Die Großmutter umarmt The grandmother hugs	die hübsche the pretty	Enkelin. granddaughter.	0.53
3b	Die Großmutter zerbricht The grandmother breaks	die hübsche the pretty	Vase. vase.	0.79
4a	Die Ärztin impft The doctor (fem.) vaccinates	den kleinen the small	Jungen. boy.	0.60
4b	Die Ärztin zerreißt The doctor (fem.) rips	den kleinen the small	Zettel. slip of paper.	0.59
5a	Die Lehrerin beruhigt The teacher (fem.) calms down	das kleine the small	Kind. child.	0.53 (0.46 for girl)
5b	Die Lehrerin parkt The teacher (fem.) parks	das kleine the small	Auto. car.	0.91
6a	Das Mädchen tröstet The girl comforts	das süße the sweet	Kind. child.	0.29 (0.24 for baby)

continued on next page

6b	Das Mädchen nascht The girl nibbles	das süße the sweet	Gebäck. pastry.	0.29
7a	Die Frau beauftragt The woman instructs	den guten the good	Handwerker. handyman.	0.26
7b	Die Frau trinkt The woman drinks	den guten the good	Wein. wine.	0.43
8a	Der Vater kitzelt The father tickles	den kleinen the small	Jungen. boy.	0.40 (0.31 for son)
8b	Der Vater steuert The father steers	den kleinen the small	Wagen. vehicle.	0.52
9a	Das Mädchen ruft The girl calls	den großen the big	Bruder. brother.	0.65
9b	Das Mädchen wirft The girl throws	den großen the big	Ball. ball.	0.54
10a	Der Junge ärgert The boy annoys	die große the big	Schwester. sister.	0.71
10b	Der Junge baut The boy builds	die große the big	Burg. castle.	0.53 (0.12 for sand castle)
11a	Der Gärtner verjagt The gardener chases away	den störenden the disruptive	Hund. dog.	0.24
11b	Der Gärtner kürzt The gardener trims	den störenden the disruptive	Busch. bush.	0.31
12a	Das Mädchen streichelt The girl pets	das schöne the beautiful	Pferd. horse.	0.29 (0.14 for pony)
12b	Das Mädchen beschmutzt The girl stains	das schöne the beautiful	Kleid. dress.	0.47
13a	Der Zauberer beneidet The magician envies	den beliebten the popular	Clown. clown.	0.14
13b	Der Zauberer probt The magician practices	den beliebten the popular	Trick. trick.	0.50
14a	Die Frau dressiert The woman trains	den hellen the pale	Hund. dog.	0.51 (other for dog: 0.20)
14b	Die Frau saugt The woman vacuums	den hellen the pale	Teppich. carpet.	0.74
15a	Die Mutter wickelt The mother swaddles	das süße the sweet	Baby. baby.	0.66
15b	Die Mutter püriert The mother blends	das süße the sweet	Obst. fruit.	0.26
16a	Der Mann bestraft The man punishes	den lauten the loud	Hund. dog.	0.35
16b	Der Mann repariert	den lauten	Motor.	0.34

continued on next page

	The man repairs	the loud	engine.	
17a	Der Arzt röntgt	den neuen	Patienten.	0.63
	The doctor x-rays	the new	patient.	
17b	Der Arzt fährt	den neuen	Wagen.	0.37
	The doctor drives	the new	vehicle.	(similar words: 0.54)
18a	Die Sekretärin informiert	den neuen	Chef.	0.49
	The secretary informs	the new	boss.	
18b	Die Sekretärin beschriftet	den neuen	Ordner.	0.37
	The secretary labels	the new	folder.	
19a	Der Student fragt	den alten	Professor.	0.54
	The student asks	the old	professor.	
19b	Der Student updatet	den alten	Computer.	0.46
	The student updates	the old	computer.	
20a	Die Diebin warnt	den versteckten	Komplizen.	0.49
	The thief (fem.) warns	the hidden	accomplice.	
20b	Die Diebin knackt	den versteckten	Tresor.	0.46
	The thief (fem.) cracks	the hidden	safe.	(synonym: 0.26)
21a	Der Bräutigam heiratet	die schöne	Braut.	0.66
	The groom marries	the beautiful	bride.	
21b	Der Bräutigam schneidet	die schöne	Torte.	0.58
	The groom slices	the beautiful	cake.	
22a	Die Mutter frisiert	die kleine	Tochter.	0.62
	The mother coiffures	the small	daughter's hair.	
22b	Die Mutter strickt	die kleine	Mütze.	0.27
	The mother knits	the small	hat.	
23a	Die Lehrerin ermahnt	den schwierigen	Schüler.	0.83
	The teacher (fem.) reprimands	the difficult	pupil.	
23b	Die Lehrerin verteilt	den schwierigen	Test.	0.74
	The teacher (fem.) distributes	the difficult	test.	
24a	Der Postbote besänftigt	den großen	Hund.	0.86
	The postman calms down	the big	dog.	
24b	Der Postbote knickt	den großen	Umschlag.	0.51
	The postman folds	the big	envelope.	(similar words: 0.37)

Table B.1: Experimental items in Experiment 1. The last column shows the cloze probability for the specific target word in the sentence. If participants in the cloze test used another word, e.g., a car brand instead of the more general word vehicle, or a word that could be displayed by a very similar picture, the cloze probability for the other word(s) is given in parentheses.

B.2 Experiment 2: Experimental items

No	**adverbial + verb**	CRITICAL WINDOW	**quantifier + noun**
1a/b	Am Baum ist/sind	tatsächlich	ein Apfel/mehrere Äpfel.
	On the tree is/are	indeed	an apple/several apples
2a/b	*Neben dem Telefon ist/sind*	*immer*	*ein Stift/einige Stifte.*
	Next to the telephone is/are	*always*	*a pen/some pens*
3a/b	*Im Kühlschrank ist/sind*	*immer*	*eine Flasche Milch/mehrere Flaschen Milch.*
	In the fridge is/are	*always*	*a bottle of milk/several bottles of milk*
4a/b	Auf der Straße ist/sind	am Mittag	ein Auto/einige Autos.
	On the street is/are	at noon	a car/some cars
5a/b	*In der Schublade ist/sind*	*immer*	*eine Briefmarke/mehrere Briefmarken.*
	In the drawer is/are	*always*	*a stamp/several stamps*
6a/b	Auf den Stufen ist/sind	häufiger	eine Katze/einige Katzen.
	On the steps is/are	often	a cat/some cats
7a/b	Im Teich ist/sind	neuerdings	ein Goldfisch/mehrere Goldfische.
	In the pond is/are	lately	a goldfish/several goldfish
8a/b	Im Nest ist/sind	tatsächlich	ein Ei/einige Eier.
	In the nest is/are	indeed	an egg/some eggs
9a/b	Vor dem Fenster ist/sind	früh morgens	ein Vogel/mehrere Vögel.
	In front of the window is/are	early in the morning	a bird/several birds
10a/b	Im Bad ist/sind	am Abend	eine Mücke/einige Mücken.
	In the bathroom is/are	in the evening	a mosquito/some mosquitoes
11a/b	Im Briefkasten ist/sind	am Mittag	eine Zeitung/mehrere Zeitungen.
	In the mailbox is/are	at noon	a newspaper/several newspapers
12a/b	Vor der Tür ist/sind	früh morgens	ein Paket/einige Pakete.
	In front of the door is/are	early in the morning	a package/some packages
13a/b	*In der Vase ist/sind*	*immer*	*eine Blume/mehrere Blumen.*
	In the vase is/are	*always*	*a flower/several flowers*
14a/b	Auf der Schaukel ist/sind	am Mittag	ein Kind/einige Kinder.
	On the swing is/are	at noon	a child/some children
15a/b	Auf dem Fluss ist/sind	häufiger	ein Boot/mehrere Boote.
	On the river is/are	often	a boat/several boats
16a/b	Am Himmel ist/sind	nachmittags	ein Helikopter/einige Helikopter.
	In the sky is/are	in the afternoon	a helicopter/some helicopters
17a/b	An der Haltestelle ist/sind	häufiger	ein Reisebus/mehrere Reisebusse.
	At the bus stop is/are	often	a coach/several coaches
18a/b	Auf dem Teich ist/sind	regelmäßig	eine Ente/einige Enten.

continued on next page

	On the pond is/are	regularly	a duck/some ducks
19a/b	Auf der Blume ist/sind	tatsächlich	ein Schmetterling/mehrere Schmetterlinge.
	On the flower is/are	indeed	a butterfly/several butterflies
20a/b	Auf dem Tisch ist/sind	am Morgen	eine Kerze/einige Kerzen.
	On the table is/are	in the morning	a candle/some candles
21a/b	Am Himmel ist/sind	am Abend	ein Stern/mehrere Sterne.
	In the sky is/are	in the evening	a star/several stars
22a/b	Auf dem Herd ist/sind	am Mittag	ein Topf/einige Töpfe.
	On the stove is/are	at noon	a pot/some pots
23a/b	An der Wand ist/sind	neuerdings	ein Bild/mehrere Bilder.
	On the wall is/are	lately	a picture/several pictures
24a/b	Hinter dem Zaun ist/sind	neuerdings	ein Pferd/einige Pferde.
	Behind the fence is/are	lately	a horse/some horses
25a/b	Auf dem Tisch ist/sind	nachmittags	eine Tasse Tee/mehrere Tassen Tee.
	On the table is/are	in the afternoon	a cup of tea/several cups of tea
26a/b	Über dem Berg ist/sind	nachmittags	eine Regenwolke/einige Regenwolken.
	Above the mountain is/are	in the afternoon	a rain cloud/some rain clouds
27a/b	Im Ofen ist/sind	Heiligabend	eine Gans/mehrere Gänse.
	In the oven is/are	on Christmas Eve	a goose/several geese
28a/b	Im Wasser ist/sind	tatsächlich	ein Pinguin/einige Pinguine.
	In the water is/are	indeed	a penguin/some penguins

Table B.2: Experimental items in Experiment 2. Sentences shown in italics were removed from the analyses because the critical window was much shorter than for all others. Below each German sentence, the word-by-word English translation is given.

B.3 Experiment 3: Experimental items

No	animacy	**NP1, Agent verb NP2**	CRITICAL WINDOW	**Theme (a), Recipient (b)**
1a	-animate	Der Gärtner gibt d**er** blühend**en**	Pflanze eilig	frisches Wasser.
		The gardener gives the-*Dat* flowering-*Dat*	plant quickly	fresh water
1b	-animate	Der Gärtner gibt d**ie** blühend**e**	Pflanze eilig	dem Postboten.
		The gardener gives the-*Acc* flowering-*Acc*	plant quickly	to the postman
	pictures: gardener, flowering plant, watering can, postman			
2a	-animate	Der Junge gibt d**er** neu**en**	Actionfigur sofort	ein Schwert.
		The boy gives the-*Dat* new-*Dat*	action figure immediately	a sword
2b	-animate	Der Junge gibt d**ie** neu**e**	Actionfigur sofort	dem Vater.
		The boy gives the-*Acc* new-*Acc*	action figure immediately	to the father
	pictures: boy, action figure, sword, man			

continued on next page

3a	+animate	Der Junge gibt d**er** griechisch**en** The boy gives the-*Dat* Greek-*Dat*	Schildkröte vorsichtig tortoise carefully	ein Salatblatt. a lettuce leaf
3b	+animate	Der Junge gibt d**ie** griechisch**e** The boy gives the-*Acc* Greek-*Acc*	Schildkröte vorsichtig tortoise carefully	dem Tierarzt. to the veterinarian
	pictures: boy, tortoise, lettuce leaf, veterinarian			
4a	-animate	Der Postbote gibt de**m** ausländisch**en** The postman gives the-*Dat* foreign-*Dat*	Paket eilig package quickly	einen Stempel. a stamp
4b	-animate	Der Postbote gibt d**as** ausländisch**e** The postman gives the-*Acc* foreign-*Acc*	Paket eilig package quickly	der Sekretärin. to the secretary
	pictures: postman, package, stamp, woman/secretary			
5a	-animate	Die Sekretärin gibt d**em** piepend**en** The secretary gives the-*Dat* beeping-*Dat*	Kopiergerät schließlich photocopier finally	neues Papier. new paper
5b	-animate	Die Sekretärin gibt d**as** piepend**e** The secretary gives the-*Acc* beeping-*Acc*	Kopiergerät schließlich photocopier	dem Reparaturdienst. to the repair service
	pictures:woman/secretary, copying machine, pile of paper, handyman			
6a	-animate	Der Sportler gibt d**er** golden**en** The sportsman gives the-*Dat* golden-*Dat*	Medaille sofort medal immediately	einen Kuss. a kiss
6b	-animate	Der Sportler gibt d**ie** golden**e** The sportsman gives the-*Acc* golden-*Acc*	Medaille sofort medal immediately	der kleinen Tochter. to the little daughter
	pictures: sportsman, medal, kissing lips, girl			
7a	-animate	Der Student gibt d**er** alt**en** The student gives the-*Dat* old-*Dat*	Waschmaschine schließlich washing machine finally	einen Tritt. a kick
7b	-animate	Der Student gibt d**ie** alt**e** The student gives the-*Acc* old-*Acc*	Waschmaschine schließlich washing machine finally	dem Handwerker. to the handyman
	pictures: young man, washing machine, kicking foot, handyman			
8a	+animate	Der Pirat gibt d**em** schreien**den** The pirate gives the-*Dat* shrieking-*Dat*	Äffchen eilig monkey quickly	eine Banane. a banana
8b	+animate	Der Pirat gibt d**as** schreiend**e** The pirate gives the-*Acc* shrieking-*Acc*	Äffchen eilig monkey quickly	dem Matrosen. to the sailor
	pictures: pirate, monkey, banana, sailor			
9a	+animate	Die Ärztin übergibt d**er** krank**en** The doctor hands the-*Dat* sick-*Dat*	Patientin schließlich patient finally	das Rezept. the prescription
9b	+animate	Die Ärztin übergibt d**ie** krank**e** The doctor hands over the-*Acc* sick-*Acc*	Patientin schließlich patient finally	dem Kollegen. to the colleague
	pictures: female doctor, older lady with walking stick, prescription, male doctor			
10a	+animate	Der Polizist übergibt d**er** bös**en** The policeman hands the-*Dat* evil-*Dat*	Diebin direkt thief (fem.) directly	den Haftbefehl. the arrest warrant
10b	+animate	Der Polizist übergibt d**ie** bös**e**	Diebin direkt	der Dienstwache.

continued on next page

		The policeman hands over the-*Acc* evil-*Acc*	thief directly	to the police station
	pictures: policeman, female thief, arrest warrant, police department			
11a	+animate	Die Lehrerin übergibt d**er** frech**en**	Schülerin wortlos	den schlechten Test.
		The teacher hands the-*Dat* nasty-*Dat*	pupil without a word	the bad test.
11b	+animate	Die Lehrerin übergibt d**ie** frech**e**	Schülerin wortlos	dem Direktor.
		The teacher hands over the-*Acc* nasty-*Acc*	pupil without a word	to the headmaster
	pictures: female teacher, girl, test, man with tie/headmaster			
12a	+animate	Der Vater übergibt d**er** glücklich**en**	Tochter direkt	ein Geschenk.
		The father hands the-*Dat* happy-*Dat*	daughter directly	a present
12b	+animate	Der Vater übergibt d**ie** glücklich**e**	Tochter direkt	dem Bräutigam.
		The father hands over the-*Acc* happy-*Acc*	daughter directly	to the groom
	pictures: older man, bride, present, groom			
13a	+animate	Die alte Dame überreicht d**er** verwöhnt**en**	Katze vorsichtig	das Futter.
		The old lady hands the-*Dat* spoiled-*Dat*	cat carefully	the food
13b	+animate	Die alte Dame überreicht d**ie** verwöhnt**e**	Katze vorsichtig	dem Neffen.
		The old lady hands the-*Acc* spoiled-*Acc*	cat carefully	to the nephew
	pictures: old lady, cat, feeding dish, young man			
14a	+animate	Der Vater überreicht d**em** schreiend**en**	Baby vorsichtig	den Schnuller.
		The father hands the-*Dat* crying-*Dat*	baby carefully	the pacifier
14b	+animate	Der Vater überreicht da**s** schreiend**e**	Baby vorsichtig	der Mutter.
		The father hands the-*Acc* crying-*Acc*	baby carefully	to the mother
	pictures: man, baby, pacifier, woman			
15a	-animate	Der Clown überreicht d**er** lustig**en**	Handpuppe tatsächlich	das Mikrofon.
		The clown hands the-*Dat* funny-*Dat*	hand puppet actually	the microphone
15b	-animate	Der Clown überreicht d**ie** lustig**e**	Handpuppe tatsächlich	der Zuschauerin.
		The clown hands the-*Acc* funny-*Acc*	hand puppet actually	to the spectator
	pictures: clown, hand puppet, microphone, woman			
16a	+animate	Die Mutter überreicht d**er** lachend**en**	Tochter sofort	den Ball.
		The mother hands the-*Dat* laughing-*Dat*	daughter immediately	the ball
16b	+animate	Die Mutter überreicht d**ie** lachend**e**	Tochter sofort	dem Vater.
		The mother hands the-*Acc* laughing-*Acc*	daughter immediately	to the father
	pictures: woman, girl, ball, man			
17a	*+animate*	*Der Zwerg überlässt d**er** bös**en***	*Hexe zögerlich*	*den Diamanten.*
		The dwarf surrenders the-Dat evil-Dat	*witch hesitantly*	*the diamond*
17b	*+animate*	*Der Zwerg überlässt d**ie** bös**e***	*Hexe zögerlich*	*dem König.*
		The dwarf surrenders the-Acc evil-Acc	*witch hesitantly*	*to the king*
	pictures: dwarf, witch, diamond, king			
18a	+animate	Der Dieb überlässt d**er** trügerisch**en**	Komplizin ungern	die Beute.

continued on next page

		The thief relinquishes the-*Dat* treacherous-*Dat*	accomplice reluctantly	the loot
18b	+animate	Der Dieb überlässt d**ie** trügerisch**e**	Komplizin ungern	dem Polizisten.
		The thief relinquishes the-*Acc* treacherous-*Acc*	accomplice reluctantly	to the policeman
	pictures: male thief, female thief, money, policeman			
19a	-animate	Die Großmutter kauft d**er** niedlich**en**	Puppe tatsächlich	neue Kleider.
		The grandmother buys the-*Dat* cute-*Dat*	doll indeed	new clothes
19b	-animate	Die Großmutter kauft d**ie** niedlich**e**	Puppe tatsächlich	dem Mädchen.
		The grandmother buys the-*Acc* cute-*Acc*	doll indeed	for the girl
	pictures: old lady, doll, dresses, girl			
20a	-animate	Die Sekretärin kauft d**er** robust**en**	Grünpflanze schließlich	einen neuen Topf.
		The secretary buys the-*Dat* robust-*Dat*	house plant finally	a new pot
20b	-animate	Die Sekretärin kauft d**ie** robust**e**	Grünpflanze schließlich	ihrem Chef.
		The secretary buys the-*Acc* robust-*Acc*	house plant finally	for her boss
	pictures: woman/secretary, plant, pot, man with tie/boss			
21a	+animate	Der Chef kauft d**em** erfolgreich**en**	Rennpferd tatsächlich	einen neuen Sattel.
		The boss buys the-*Dat* successful-*Dat*	race horse indeed	a new saddle
21b	+animate	Der Chef kauft d**as** erfolgreich**e**	Rennpferd tatsächlich	dem Sohn.
		The boss buys the-*Acc* successful-*Acc*	race horse indeed	for the son
	pictures: man with tie/boss, horse, saddle, young man			
22a	*+animate*	*Der König verspricht d**er** schön**en***	*Prinzessin tatsächlich*	*die Krone.*
		The king promises the-Dat beautiful-Dat	*princess indeed*	*the crown*
22b	*+animate*	*Der König verspricht d**ie** schön**e***	*Prinzessin tatsächlich*	*dem siegreichen Ritter*
		The king promises the-Acc beautiful-Acc	*princess indeed*	*to the victorious knight*
	pictures: king, princess, crown, knight			
23a	*+animate*	*Die Ärztin zeigt d**em** neugeboren**en***	*Baby sofort*	*den Teddy.*
		The doctor shows the-Dat newborn-Dat	*baby immediately*	*the teddy*
23b	*+animate*	*Die Ärztin zeigt d**as** neugeboren**e***	*Baby sofort*	*dem Vater.*
		The doctor shows the-Acc newborn-Acc	*baby immediately*	*to the father*
	pictures: female doctor, baby, teddy bear, man			
24a	-animate	Das Mädchen zeigt d**em** weich**en**	Kuscheltier sofort	das Bilderbuch.
		The girl shows the-*Dat* soft-*Dat*	stuffed animal immediately	the picture book
24b	-animate	Das Mädchen zeigt d**as** weich**e**	Kuscheltier sofort	dem Vater.
		The girl shows the-*Acc* soft-*Acc*	stuffed animal immediately	to the father
	pictures: girl, stuffed animal, picture book, man			
25a	+animate	Der Forscher zeigt d**em** intelligente**n**	Äffchen erneut	eine Symbol-Abfolge.
		The researcher shows the-*Dat* intelligent-*Dat*	monkey again	a sequence of symbols
25b	+animate	Der Forscher zeigt da**s** intelligent**e**	Äffchen erneut	der Kollegin.
		The researcher shows the-*Acc* intelligent-*Acc*	monkey again	to the colleague

continued on next page

	pictures: female researcher, monkey, plate with symbols, male researcher			
26a	*+animate*	*Der Chef präsentiert d**er** neu**en*** *The boss presents the-Dat new-Dat*	*Kollegin direkt* *colleague directly*	*den Arbeitsplatz.* *the workplace*
26b	*+animate*	*Der Chef präsentiert d**ie** neu**e*** *The boss presents the-Acc new-Acc*	*Kollegin direkt* *colleague directly*	*den anderen Mitarbeitern.* *to the other staff members*
	pictures: man with tie/boss, woman, computer/screen, group of people			
27a	*+animate*	*Der Student präsentiert d**er** aktuell**en*** *The student presents the-Dat current-Dat*	*Freundin schließlich* *girlfriend finally*	*sein Motorrad.* *his motorbike*
27b	*+animate*	*Der Student präsentiert d**ie** aktuell**e*** *The student presents the-Acc current-Acc*	*Freundin schließlich* *girlfriend finally*	*den Eltern.* *to his parents*
	pictures: young man, young woman, motorbike, couple			
28a	+animate	Der Zauberer präsentiert d**em** weiß**en** The magician presents the-*Dat* white-*Dat*	Kaninchen eilig rabbit quickly	eine Karotte. a carrot
28b	+animate	Der Zauberer präsentiert d**as** weiß**e** The magician presents the-*Acc* white-*Acc*	Kaninchen eilig rabbit quickly	dem Publikum. to the audience
	pictures: magician, white rabbit, carrot, audience			

Table B.3: Experimental items in Experiment 3. Condition (a) follows the canonical order 'dative object/Recipient precedes accusative object/Theme,' condition (b) the non-canonical order 'accusative object/Theme precedes dative object/Recipient.' Those sentences that were excluded because the competitor in the (a) condition would have been a plausible continuation are shown in italics. Below the German sentences, the English word-by-word translation is given. The second column indicates whether the first post-verbal argument was counted as +animate (human, animal) or -animate (plant, object) in the post-hoc analysis.

B.4 Experiment 1–3: Filler items

No	Set	
1	number	Auf dem Baum sitzt mittags eine Schar Vögel. 'At noon, a flight of birds sits in the tree.'
	pictures: tree, single bird, birds	
2	number	Auf der Straße spielt am Abend eine Gruppe Kinder. 'In the evening, a group of children plays in the street.'
	pictures: street, single child, children	
3	number	Auf dem Tisch steht morgens ein Haufen Gläser. 'In the morning, a lot of glasses are standing on the table.'
	pictures: table, single glass, glasses	
4	number	Im Park ist neuerdings eine Bande Diebe. 'In the park, there is recently a gang of thieves.'
	pictures: park, single thief, thieves	

continued on next page

5	number	In der Kirche weinen Heiligabend mehrere Babys. 'On Christmas Eve, several babies in the church are crying.'
	pictures: church, single baby, babies	
6	number	Auf dem Sofa liegen immer einige Decken. 'On the sofa, there are always some blankets.'
	pictures: sofa, single blanket, blankets	
7	number	Im Park stehen neuerdings mehrere Mülltonnen. 'In the park, there are several trash bins recently.'
	pictures: park, single trash bin, trash bins	
8	number	Im Unterricht klingelt häufiger ein Smartphone. 'During a class, a smartphone often rings.'
	pictures: blackboard, single smartphone, smartphones	
9	number	Hinter dem Haus bellt am Abend ein Hund. 'Behind the house, a dog barks in the evening.'
	pictures: house, single dog, dogs	
10	number	Im Bad hängt immer ein Handtuch. 'In the bathroom, there is always a towel.'
	pictures: bathroom, single towel, towels	
11	mixed	Auf dem Berg wachsen tatsächlich einige Bäume. 'On the mountain, there are indeed growing some trees.'
	pictures: mountain, single tree, trees, house	
12	mixed	Durch den Park spazieren täglich viele Menschen. 'Many people walk through the park each day.'
	pictures: park, single person, several people, cat	
13	mixed	Am Teich sind am Abend viele Mücken. 'In the evening, there are many mosquitoes around the pond.'
	pictures: pond, single mosquito, mosquitoes, bird	
14	mixed	Der Junge isst eine handvoll Kirschen. 'The boy eats a handful of cherries.'
	pictures: boy, single cherry, cherries, sand castle	
15	mixed	Der Forscher trinkt vormittags mehrere Tassen Kaffee. 'The researcher drinks several cups of coffee before noon.'
	pictures: researcher, single cup, cups, cake	
16	mixed	Der Polizist verhaftet schließlich die Diebe. 'The policeman finally arrests the thieves.'
	pictures: policeman, single thief, thieves, safe	
17	case	Der Postbote überreicht den erwarteten Brief der Mutter. 'The postman hands the expected letter to the mother.'

continued on next page

	pictures: postman, envelope, woman	
18	case	Die Großmutter gibt die Zeitung gleich dem Vater. 'The grandmother promptly gives the father the newspaper.'
	pictures: old lady, newspaper, man, package	
19	case	Den selbstgebackenen Kuchen bringt die kleine Tochter stolz der Mutter. 'The little daughter proudly brings the home-made cake to the mother.'
	pictures: girl, cake, woman	
20	case	Der Zauberer zeigt den schweren Trick ungern den Freunden. 'The magician reluctantly shows the difficult trick to his friends.'
	pictures: magician, cards, group of people, white rabbit	
21	case	Der Bräutigam übergibt den Ring der glücklichen Braut. 'The groom hands the ring to his happy bride.'
	pictures: bridegroom, ring, bride	
22	case	Der Ritter bringt die böse Hexe dem König. 'The knight brings the evil witch to the king.'
	pictures: knight, witch, king, diamond	
23	case	Vor der Dienstwache flüchtet mit dem Motorrad die Diebin. 'The thief (fem.) escapes with the motorbike in front of the police station.'
	pictures: female thief, motorbike, police station	
24	case	Das Fenster der Großmutter repariert am Abend der Handwerker. 'In the evening, the handyman repairs the grandmother's window.'
	pictures: old lady, window, handyman, door	
25	case	Den schweren Ordner gibt die Sekretärin dem Forscher. 'The secretary gives the heavy folder to the researcher'.
	pictures: woman/secretary, folder, researcher	
26	case	Der Mitarbeiter stellt auf den Tisch der Sekretärin den Kaffee. 'The staff member puts the coffee on the secretary's table.'
	pictures: man, table, woman/secretary, cup	
27	case	Die Lehrerin nimmt dem Schüler das Smartphone weg. 'The teacher takes away the student's smartphone.'
	pictures: female teacher, student, smartphone	
28	case	Der Zwerg schenkt den Diamanten der schönen Prinzessin. 'The dwarf presents the diamond to the beautiful princess.'
	pictures: dwarf, diamond, princess, witch	

Table B.4: Filler items for Experiments 1–3. The second column indicates which item set the fillers resembled.

B.5 Instructions for participants

Willkommen zum Experiment!

In diesem Experiment wirst du immer einige Bilder sehen und dazu einen Satz hören. Bitte höre aufmerksam zu! Nach einigen dieser Sätze wird ein geschriebener Satz auf dem Bildschirm erscheinen. Deine Aufgabe ist es, zu entscheiden, ob dieser Satz zu dem zuvor Gehörten passt. Entscheidest du dich für „Nein“ (stimmt nicht), drücke die linke Schultertaste auf dem Gamepad, entscheidest du dich für „Ja“ (stimmt), drücke die rechte Schultertaste. Es geht dabei nicht um Schnelligkeit, sondern darum, dass deine Antwort möglichst korrekt ist, du darfst dir also ruhig etwas Zeit lassen.

Wir beginnen mit einem kurzen Übungsblock. Danach hast du die Möglichkeit, Fragen zu stellen. Es wird zwei Pausen geben, in denen du dich gern bewegen kannst. Während des Experiments wäre es gut, wenn du dich so wenig wie möglich bewegst, damit die Kamera deine Blickbewegungen aufzeichnen kann. Um zu kontrollieren, dass die Kamera deine Blickbewegungen richtig erfasst, starten wir mit einer sog. Kalibrierung, diese wird nach den Pausen wiederholt. Dabei folgst du mit deinen Blicken einem Punkt auf dem Bildschirm, der sich bewegen wird.

English translation:

Welcome to the experiment!

In this experiment you are going to see several pictures while you listen to a sentence. Please listen carefully! After some of these sentences a written sentence will appear on the screen. Your task is to decide whether this sentence matches what you have heard. If you decide for "No" (does not match), press the left shoulder key on your gamepad, if you decide for "Yes" (match), press the right shoulder key. It is not important that you respond as fast as possible but rather that you respond accurately, so you can take your time.

We will start with a short practice block. Afterwards you have the chance to ask questions. There will be two breaks, during which you can move around as you like. During the experiment it would be good if you move as little as possible, so the camera can track your eye-movements. To check that the camera is tracking your eye-movements accurately, we will start with a so-called calibration, which will be repeated after the breaks. For this you need to follow a moving dot on the screen with your eyes.

B.6 Additional tables

	Estimate	Std. Error	*t*-value	*p*-value
Intercept (L1, singular)	-0.3904	0.2381	-1.640	0.106
Condition (plural)	-0.2851	0.3770	-0.756	0.452
Group (L2)	-0.1762	0.3293	-0.535	0.594
Linear time	0.4482	0.4122	1.087	0.282
Condition x Group	0.3613	0.5249	0.688	0.494
Condition x Linear time	-0.6674	0.6017	-1.109	0.273
Group x Linear time	0.4253	0.5982	0.711	0.481
Condition x Group x Linear time	-0.5850	0.8685	-0.674	0.504
Intercept (L2, singular)	-0.5667	0.2401	-2.360	0.0214 *
Condition (plural)	0.0761	0.3711	0.205	0.8381
Linear time	0.8735	0.4335	2.015	0.0495 *
Condition x Linear time	-1.2524	0.6263	-2.000	0.0511 .
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				
Formula: Elog ~ condition*group*ot1 + (1 + condition*group | item) + (1 + condition*ot1 | subject)				

Table B.5: Results of the growth curve analyses for the critical window in Experiment 2 with treatment contrast coding for condition and group

	Estimate	Std. Error	*t*-value	*p*-value
Intercept (L1, canonical)	0.0692	0.2487	0.278	0.7823
Condition (non-canonical)	-0.1742	0.3069	-0.567	0.5726
Group (L2)	-0.2908	0.295	-0.986	0.3286
Linear time	2.9173	0.8164	3.574	0.0008 ***
Quadratic time	0.532	0.6127	0.868	0.3893
Condition x Group	0.4831	0.4157	1.162	0.2498
Condition x Linear time	-2.2393	1.1703	-1.913	0.0617 .
Condition x Quadratic time	-1.3793	0.8659	-1.593	0.1174
Group x Linear time	-0.2121	1.195	-0.178	0.8598
Group x Quadratic time	0.1418	0.891	0.159	0.8742
Condition x Group x Linear time	0.4047	1.6998	0.238	0.8128
Condition x Group x Quadratic time	-0.3766	1.2543	-0.300	0.7652
Intercept (L1, non-canonical)	-0.105	0.2273	-0.462	0.6458
Group (L2)	0.1924	0.2947	0.653	0.5164
Linear Time	0.678	0.914	0.742	0.4619
Quadratic Time	-0.8473	0.5473	-1.548	0.1284
Group x Linear Time	0.1926	1.3198	0.146	0.8846
Group x Quadratic Time	-0.2348	0.7874	-0.298	0.7669
Intercept (L2, canonical)	-0.2216	0.2140	-1.036	0.3048
Condition (non-canonical)	0.3090	0.3007	1.028	0.3083
Linear time	2.7052	0.8730	3.099	0.0032 **
Quadratic time	0.6738	0.6470	1.041	0.3027
Condition x Linear time	-1.8346	1.2332	-1.488	0.1435
Condition x Quadratic time	-1.7559	0.9077	-1.934	0.0588 .
Intercept (L2, non-canonical)	0.0874	0.2317	0.377	0.7075
Linear Time	0.8706	0.9524	0.914	0.3655
Quadratic Time	-1.0821	0.5663	-1.911	0.0626 .

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Formula: Elog ~ condition*group*(ot1+ot2) + (1 + condition*group | item) + (1 + condition*(ot1+ot2) | subject)

Table B.6: Results of the growth curve analyses for the critical window in Experiment 3 with treatment contrast coding for condition and group

Appendix C

Experiment 4: Materials and additional information

C.1 Materials

C.1.1 Experimental items

The design of the experimental items followed the scheme shown in Table C.1. Each item appeared in a version (a) and a version (b), that is, the reversed order of characters as in (a), to counterbalance potential effects of social status etc. The (a) versions were tested in Pre-test 2 and are listed in Table C.2. The respective target picture is highlighted in bold, that is, for item (1a) below the target AOI corresponds to the picture of a Native American holding a knife (competitor: cowboy with rifle) and for item (1b) to the cowboy holding a knife, yielding a completely balanced design. Items with odd numbers include an NP1-biasing verb and items with even numbers an NP2-biasing verb.

No	NP1 verb NP2 distractor	CRITICAL WINDOW	disambiguating information
1a	Der **Indianer** ängstigt den Cowboy am Lagerfeuer,	weil er zufällig	ein **Messer** hervorholt.
	The **Native American** frightens the cowboy at the campfire	because he accidentally	takes out a **knife**.
1b	Der **Cowboy** ängstigt den Indianer am Lagerfeuer,	weil er zufällig	ein **Messer** hervorholt.
	The **cowboy** frightens the Native American at the campfire	because he accidentally	takes out a **knife**.
2a	Der Indianer fürchtet den **Cowboy** am Lagerfeuer,	weil er zufällig	ein **Gewehr** hervorholt.
	The Native American fears the **cowboy** at the campfire	because he accidentally	takes out a **rifle**.
2b	Der Cowboy fürchtet den **Indianer** am Lagerfeuer,	weil er zufällig	ein **Gewehr** hervorholt.
	The cowboy fears the **Native American** at the campfire	because he accidentally	takes out a **rifle**.

Table C.1: Schematic design of the experimental items in Experiment 4

No	**NP1 verb NP2 distractor**	CRITICAL WINDOW	**disambiguating information**
3a	Der **Einbrecher** verblüfft den Polizisten im Museum, The **burglar** baffles the policeman in the museum	weil er unerwartet because he unexpectedly	eine **Taschenlampe** dabeihat. has a **flashlight** with him.
4a	Der Einbrecher fürchtet den **Polizisten** im Museum, The burglar fears the **policeman** in the museum	weil er unerwartet because he unexpectedly	eine **Pistole** dabeihat. has a **gun** with him.
5a	Der **Römer** erfreut den Ägypter in der Pyramide, The **Roman** delights the Egyptian in the pyramid	weil er neuerdings because he recently	eine **Karte** dabeihat. has a **map** with him.
6a	Der Römer bevorzugt den **Ägypter** in der Pyramide, The Roman prefers the **Egyptian** in the pyramid	weil er neuerdings because he recently	eine **Laterne** dabeihat. has a **lantern** with him.
7a	*Der **Bauer** erstaunt den Gärtner in der Scheune,* *The **farmer** astonishes the gardener in the barn*	*weil er tatsächlich* *because he actually*	*einen **Kürbis** transportiert.* *moves a **pumpkin**.*
8a	*Der Bauer bemitleidet den **Gärtner** in der Scheune,* *The farmer pities the **gardener** in the barn*	*weil er tatsächlich* *because he actually*	*einen **Kornsack** transportiert.* *moves a **sack of grain**.*
9a	Der **Musiker** verblüfft den Clown auf der Bühne, The **musician** baffles the clown on the stage	weil er neuerdings because he recently	ein **Saxophon** besitzt. owns a **saxophone**.
10a	Der Musiker liebt den **Clown** auf der Bühne, The musician loves the **clown** on the stage	weil er neuerdings because he recently	ein **Schlagzeug** besitzt. owns a **drum set**.
11a	Der **Feuerwehrmann** verwirrt den Bauarbeiter auf der Straße, The **firefighter** confuses the construction worker on the street	weil er zufällig because he incidentally	einen **Feuerlöscher** bei sich trägt. carries a **fire extinguisher**.
12a	Der Feuerwehrmann liebt den **Bauarbeiter** auf der Straße, The firefighter loves the **construction worker** on the street	weil er zufällig because he incidentally	einen **Werkzeugkasten** bei sich trägt. carries a **toolbox**.
13a	Der **Taucher** empört den Piraten in der Schatzkammer, The **diver** outrages the pirate in the treasure chamber	weil er unbemerkt because he secretly	einen **Leuchter aus Gold** mitnimmt. takes a **candlestick of gold**.
14a	Der Taucher verachtet den **Piraten** in der Schatzkammer, The diver despises the **pirate** in the treasure chamber	weil er unbemerkt because he secretly	einen **Beutel Münzen** mitnimmt. takes a **bag of coins**.
15a	Der **Wächter** erschreckt den Ritter auf der Burgmauer, The **guard** scares the knight on the castle walls	weil er unerwartet because he unexpectedly	einen **Schlüssel** hervorholt. takes out a **key**.
16a	Der Wächter hasst den **Ritter** auf der Burgmauer, The guard hates the **knight** on the castle walls	weil er unerwartet because he unexpectedly	einen **Bogen** hervorholt. takes out a **bow**.
17a	Der **Verkäufer** empört den Müllmann an der Imbissbude, The **salesman** outrages the garbageman at the snack stand	weil er unbemerkt because he secretly	einen **Besen** wegnimmt. removes a **broom**.

continued on next page

18a	Der Verkäufer verachtet den **Müllmann** an der Imbiss-bude, The salesman despises the **garbageman** at the snack stand	weil er unbemerkt because he secretly	einen **Eimer** wegnimmt. removes a **bucket**.
19a	Der **Pilot** erfreut den Arzt an Weihnachten, The **pilot** delights the doctor at Christmas	weil er tatsächlich because he indeed	eine **Torte** vorbeibringt. brings a **cake**.
20a	Der Pilot verehrt den **Arzt** an Weihnachten, The pilot worships the **doctor** at Christmas	weil er tatsächlich because he indeed	eine **Tasse Kaffee** vorbeibringt. brings a **cup of coffee**.
21a	Der **Prinz** beeindruckt den König bei der Statue, The **princ**e impresses the king at the statue	weil er neuerdings because he lately	ein **Schwert** bei sich trägt. carries a **sword** with him.
22a	Der Prinz vergöttert den **König** bei der Statue, The prince adores the **king** at the statue	weil er neuerdings because he lately	ein **Zepter** bei sich trägt. carries a **sceptre** with him.
23a	Der **Cowboy** beunruhigt den Indianer in der Wüste, The **cowboy** agitates the Native American in the desert	weil er regelmäßig because he regularly	ein **Messer** bei sich trägt. carries a **knife**.
24a	Der Cowboy bemitleidet den **Indianer** in der Wüste, The cowboy pities the **Native American** in the desert	weil er regelmäßig because he regularly	ein **Gewehr** bei sich trägt. carries a **rifle**.
25a	Der **Polizist** erschreckt den Einbrecher um Mitter-nacht, The **policeman** scares the burglar at midnight	weil er tatsächlich because he indeed	eine **Taschenlampe** bei sich trägt. carries a **flashlight** with him.
26a	Der Polizist hasst den **Einbrecher** um Mitternacht, The policeman hates the **burglar** at midnight	weil er tatsächlich because he indeed	eine **Pistole** bei sich trägt. carries a **gun** with him.
27a	Der **Ägypter** erstaunt den Römer in der Ruine, The **Egyptian** astonishes the Roman in the ruin	weil er zufällig because he incidentally	eine **Karte** bei sich trägt. carries a **map** with him.
28a	Der Ägypter mag den **Römer** in der Ruine, The Egyptian likes the **Roman** in the ruin	weil er zufällig because he incidentally	eine **Laterne** bei sich trägt. carries a **lantern** with him.
29a	*Der **Gärtner** beeindruckt den Bauern auf den Feldern,* *The **gardener** impresses the farmer in the fields*	*weil er regelmäßig* *because he regularly*	*einen **Kürbis** heimbringt.* *brings home a **pumpkin**.*
30a	*Der Gärtner bewundert den **Bauern** auf den Feldern,* *The gardener admires the **farmer** in the fields*	*weil er regelmäßig* *because he regularly*	*einen **Kornsack** heimbringt.* *brings home a **sack of grain**.*
31a	Der **Clown** überrascht den Musiker im Festzelt, The **clown** surprises the musician in the party tent	weil er unerwartet because he unexpectedly	ein **Saxopho**n dabeihat. has a **saxophone** with him.
32a	Der Clown vergöttert den **Musiker** im Festzelt, The clown adores the **musician** in the party tent	weil er unerwartet because he unexpectedly	ein **Schlagzeug** dabeihat. has a **drum set** with him.
33a	Der **Bauarbeiter** begeistert den Feuerwehrmann bei der Hütte, The **construction worker** inspires the firefighter at the hut	weil er tatsächlich because he actually	einen **Feuerlöscher** mitbringt. brings along a **fire extinguisher**.
34a	Der Bauarbeiter verehrt den **Feuerwehrmann** bei der Hütte,	weil er tatsächlich	einen **Werkzeugkaste**n mit-bringt.

continued on next page

	The construction worker worships the **firefighter** at the hut	because he actually	brings along a **toolbox**.
35a	Der **Pirat** verärgert den Taucher bei den Felsen, The **pirate** annoys the diver at the rocks	weil er zufällig because he accidentally	einen **Leuchter aus Gold** findet. finds a **candlestick of gold**.
36a	Der Pirat respektiert den **Taucher** bei den Felsen, The pirate respects the **diver** at the rocks	weil er zufällig because he accidentally	einen **Beutel Münzen** findet. finds a **bag of coins**.
37a	Der **Ritter** verärgert den Wächter vor dem Gittertor, The **knight** annoys the guard in front of the gate	weil er tatsächlich because he indeed	einen **Schlüssel** bei sich trägt. carries a **key** with him.
38a	Der Ritter bedauert den **Wächter** vor dem Gittertor, The knight feels sorry for the **guard** in front of the gate	weil er tatsächlich because he indeed	einen **Bogen** bei sich trägt. carries a **bow** with him.
39a	Der **Müllmann** überrascht den Verkäufer am Brunnenbecken, The **garbageman** surprises the salesman at the fountain	weil er neuerdings because he lately	einen **Besen** dabeihat. has a **broom** with him.
40a	Der Müllmann bevorzugt den **Verkäufer** am Brunnenbecken, The garbageman prefers the **salesman** at the fountain	weil er neuerdings because he lately	einen **Eimer** dabeihat. has a **bucket** with him.
41a	Der **Arzt** verwirrt den Piloten im Krankenhaus, The **doctor** confuses the pilot in the hospital	weil er unerwartet because he unexpectedly	eine **Torte** serviert. serves a **cake**.
42a	Der Arzt mag den **Piloten** im Krankenhaus, The doctor likes the **pilot** in the hospital	weil er unerwartet because he unexpectedly	eine **Tasse Kaffee** serviert. serves a **cup of coffee**.
43a	Der **König** begeistert den Prinzen bei der Kanone, The **king** inspires the prince at the cannon	weil er tatsächlich because he actually	ein **Schwert** dabeihat. has a **sword** with him.
44a	Der König bewundert den **Prinzen** bei der Kanone, The king admires the **prince** at the cannon	weil er tatsächlich because he actually	ein **Zepter** dabeihat. has a **sceptre** with him.

Table C.2: Experimental items in Experiment 4. The items in italics indicate those excluded from the analyses of the eye-tracking experiment because of reported difficulties in picture discrimination.

C.1.2 Filler items

No	antecedent	
1	NP1	Die Jungen unterstützen den Gärtner auf der Blumenwiese, indem **sie** einen großen Korb bringen. 'The boys support the gardener in the meadow by bringing a big basket.'
2	NP2	Die Polizistin ruft beim Unfall die Feuerwehrleute, weil **sie** zum Glück einen Werkzeugkasten dabeihaben. 'The policewoman calls the firefighters at the accident, because they luckily have a toolbox with them.'
3	NP2	Die Schwimmerin begegnet dem Schwimmer an der Ladenkasse, als **er** eine Tube Sonnencreme besorgt. 'The swimmer (fem.) meets the swimmer at the till, while he is buying a tube of sunscreen.'
4	NP1	Die Polizistin entdeckt am Fenster die Einbrecher, weil **sie** aufmerksam durch ein Fernglas blickt.

continued on next page

		‘The policewoman discovers the burglars at the window, because she looks attentively through the binoculars.’
5	NP2	Die Bäuerin beobachtet die Clowns am Stehtisch, weil **sie** einen schönen Blumenstrauß halten. ‘The farmer (fem.) watches the clowns at the bar table, because they are holding a lovely bouquet of flowers.’
6	NP2	Der Kapitän erkennt die Piraten beim Leuchtturm, weil **sie** einen Säbel bei sich tragen. ‘The captain recognizes the pirates at the lighthouse, because they are holding a saber.’
7	NP1	Die Bäuerin hilft dem Bauern am Zaun, indem **sie** schnell einen Hammer herbeiholt. ‘The farmer (fem.) helps the farmer at the fence by quickly bringing a hammer.’
8	NP2	Der Schwimmer betrachtet neugierig die Kinder am Strand, weil **sie** ein Spielzeugboot haben. ‘The swimmer curiously watches the children at the beach, because they have a toy boat.’
9	NP2	Der Prinz dankt der Prinzessin im Palast, weil **sie** unerwartet den Spiegel wiederfindet. ‘The prince thanks the princess in the palace, because she unexpectedly discovers the (missing) mirror.’
10	NP2	Die Sanitäterin trifft vor der Tür die Polizisten, weil **sie** aufgeregt einen Erste-Hilfe-Koffer holen. ‘The paramedic (fem.) meets the police officers in front of the door, because they are agitatedly sending for a first-aid-kid.’
11	NP1	Die Frau grüßt freundlich den Cowboy an der Straßenlaterne, weil **sie** in der Nähe ihr Pferd sattelt. ‘The woman cordially greets the cowboy at the street lantern while saddling her horse nearby.’
12	NP1	Die Indianer sehen den Sheriff am Wasserfall, weil **sie** zufällig mit einem Hund vorbeikommen. ‘The Native Americans see the sheriff at the waterfall, because they coincidentally are walking past with a dog.’
13	NP1	Die Fee begegnet dem Zauberer auf der Blumenwiese, als **sie** einen neuen Zauberstab ausprobiert. ‘The fairy meets the wizard in the meadow while she is trying out a new magic wand.’
14	NP2	Die Reporterin beobachtet beim Autounfall den Müllmann, als **er** gerade eine Kamera hervorholt. ‘The reporter (fem.) watches the garbageman at the accident while he is unpacking a camera.’
15	NP1	Der Taucher erkennt die Tierschützerin an der Ladenkasse, als **er** schnell noch eine Packung Saft kauft. ‘The diver recognizes the animal-rights activist (fem.) at the till while he quickly buys a container of juice.’
16	NP1	Die Wirtin entdeckt die Einbrecher am Fenster, als **sie** eine Taschenlampe einschaltet. ‘The innkeeper (fem.) discovers the burglars at the window while she turns on a flashlight.’
17	NP1	Die Verkäuferin grüßt freundlich den Bauarbeiter am Stehtisch, als **sie** den Besen abstellt. ‘The saleslady cordially greets the construction worker at the bar table, while she puts aside the broom.’
18	NP2	Die Kinder betrachten neugierig den Schwimmer am Leuchtturm, als **er** ein Spielzeugboot hervorholt. ‘The children curiously watch the swimmer at the lighthouse, while he unpacks a toy boat.’
19	NP1	Die Einbrecher sehen am Zaun die Polizistin, als **sie** durch ein Fernglas blicken. ‘The burglars see the policewoman at the fence while looking through the binoculars.’
20	NP2	Die Schwimmerin ruft den Schwimmer am Strand, als **er** eine Tube Sonnencreme auspackt. ‘The swimmer (fem.) calls the swimmer on the beach, while he unpacks a tube of sunscreen.’
21	NP1	Der Kapitän unterstützt die Piraten im Palast, als **er** mit einem Säbel vorbeikommt.

continued on next page

		'The captain supports the pirates in the palace, as he comes by with a saber.'
22	NP2	Die Ärzte treffen vor der Tür die Sanitäterin, als **sie** mit einem Erste-Hilfe-Koffer vorbeigeht. 'The doctors meet the paramedic (fem.) in front of the door, while she walks by with a first-aid-kit.'
23	NP2	Der Musiker an der Straßenlaterne dankt der Verkäuferin, als **sie** eine Tasse Kaffe vorbeibringt. 'The musician thanks the saleslady, as she brings by a cup of coffee.'
24	NP1	Der Polizist hilft den Tierschützern am Wasserfall, als **er** zufällig ein Fernglas dabeihat. 'The policeman helps the animal-rights activists at the waterfall, as he coincidentally has a pair of binoculars with him.'
25	NP1	Die Bäuerin trifft auf der Blumenwiese die Clowns, während **sie** einen Blumenstrauß pflückt. 'The farmer (fem.) meets the clowns in the meadow while she picks a bouquet of flowers.'
26	NP2	Die Polizisten unterstützen die Sanitäterin beim Autounfall, während **sie** den Erste-Hilfe-Koffer holt. 'The police officers support the paramedic (fem.) at the car accident, while she retrieves the first-aid-kit.'
27	NP2	Der Arzt grüßt freundlich die Reporterin an der Ladenkasse, während **sie** ein Buch bezahlt. 'The doctor cordially greets the reporter (fem.) at the till, while she pays for a book.'
28	NP1	Der Gärtner ruft die Jungen am Fenster, während **er** einen großen Korb abstellt. 'The gardener calls the boys at the window, while he puts down a big basket.'
29	NP2	Die Verkäuferin erkennt den Musiker am Stehtisch, während **er** eine Tasse Kaffee einschenkt. 'The saleslady recognizes the musician at the bar table, while he pours a cup of coffee.'
30	NP1	Die Tierschützer entdecken den Polizisten beim Leuchtturm, während **sie** durch ein Fernglas blicken. 'The animal-rights activists discover the policeman at the lighthouse while looking through the binoculars.'
31	NP2	Die Frau dankt dem Cowboy am Zaun, während **er** das Pferd sattelt. 'The woman thanks the cowboy at the fence, while he saddles the horse.'
32	NP1	Der Verkäufer begegnet am Strand der Tierschützerin, während **er** einen Apfel isst. 'The salesman meets the animal-rights activist (fem.) at the beach, while he eats an apple.'
33	NP1	Der Ritter beobachtet im Palast die Prinzessinnen, während **er** mit dem Hund spielt. 'The knight watches the princesses in the palace while he plays with the dog.'
34	NP1	Der Prinz sieht die Prinzessin vor der Tür, während **er** in einen Spiegel schaut. 'The prince sees the princess in front of the doors while he looks into a mirror.'
35	NP2	Der Müllmann betrachtet neugierig die Reporterin an der Straßenlaterne, während **sie** eine Kamera auspackt. 'The garbageman curiously watches the reporter (fem.) at the street lantern, while she unpacks a camera.'
36	NP2	Der Zauberer hilft der Fee am Wasserfall, während **sie** offensichtlich ihren Zauberstab benutzt. 'The wizard helps the fairy at the waterfall, while she is blatantly using her magic wand.'
37	NP1	Die Prinzessinnen erkennen den Ritter auf der Blumenwiese, obwohl **sie** gerade mit dem Hund spielen. 'The princesses recognize the knight in the meadow although they are currently playing with the dog.'
38	NP1	Die Ärzte rufen die Sanitäterin beim Autounfall, obwohl **sie** den Erste-Hilfe-Koffer haben. 'The doctors call the paramedic (fem.) at the car accident although they have the first-aid-kit.'
39	NP2	Der Verkäufer an der Ladenkasse grüßt freundlich die Tierschützerin, obwohl **sie** lediglich einen Apfel kauft.

continued on next page

		'The salesman cordially greets the animal-rights activist (fem.) although she only buys one apple.'
40	NP2	Die Reporterin beobachtet neugierig den Arzt am Stehtisch, obwohl **er** lediglich ein Buch liest. 'The reporter (fem.) curiously watches the doctor at the bar table, although he is just reading a book.'
41	NP2	Der Kapitän unterstützt im Palast die Wirtin, obwohl **sie** ein vergoldetes Messer mitnimmt. 'The captain supports the innkeeper (fem.) in the palace although she steals a golden knife.'
42	NP1	Die Einbrecher begegnen vor der Tür der Wirtin, obwohl **sie** eine Taschenlampe dabeihaben. 'The burglars meet the innkeeper (fem.) in front of the door although they have a flashlight with them.'
43	NP1	Die Polizistin hilft den Feuerwehrleuten an der Straßenlaterne, obwohl **sie** nur einen Werkzeugkasten dabeihat. 'The policewoman helps the fire fighters at the street lantern although she only has a toolbox with her.'
44	NP2	Die Indianer treffen am Wasserfall den Sheriff, obwohl **er** selten dort mit seinem Hund spazieren geht. 'The Native Americans meet the sheriff at the waterfall, although he seldom walks by there with his dog.'

Table C.3: Filler items for Experiment 4. The second column indicates whether the antecedent of the pronoun was the NP1 or NP2, which was counterbalanced. Similar to the experimental items, all filler items mentioned a location and an object or instrument.

C.1.3 Instructions for participants

Im folgenden Experiment wirst du immer eine kurze Geschichte hören, während du auf dem Bildschirm drei Bilder siehst. Danach erscheint ein neuer Bildschirm, der jetzt nur noch zwei Bilder zeigt, und dir wird eine Frage zu der eben gehörten Geschichte gestellt. Deine Aufgabe ist es, diese Frage anhand der Bilder zu beantworten. Dafür drückst du die **linke Schultertaste** auf dem Controller **für das linke Bild** und die **rechte Schultertaste** auf dem Controller **für das rechte Bild**. Es geht dabei nicht um Schnelligkeit, versuche aufmerksam zuzuhören und die Fragen möglichst korrekt zu beantworten! Du kannst selbst bestimmen, wann du weitermachen magst, indem du **nach der Beantwortung der Frage** die **Taste A** drückst.

Es wird eine feste Pause geben, in der du dich gern bewegen kannst, versuche während des Experiments ansonsten bitte möglichst ruhig zu sitzen und nicht den Kopf zu bewegen! Wir starten mit einem Übungsblock, solltest du danach noch Fragen haben, kannst du sie gern stellen. Viel Spaß!

English translation:
In the following experiment you are going to hear a short story, while seeing three pictures on the screen. Afterwards, a new screen will appear showing only two pictures and you will hear a question regarding this story. Your task is to answer this question using the pictures. For this, you press the **left shoulder key** on the gamepad **for the left picture** and the **right shoulder key** on the gamepad **for the right picture**. This is not about speed; please listen carefully and try to answer as accurately as possible! You decide when you would like to continue **after you have given the response** by pressing the **button A**.

There is one set break during which you can move around as you like; please try to sit still and do not move your head during the experiment! We will start with a practice block; if you have any questions afterwards, feel free to ask. Have fun!

C.2 Additional tables

NP1-biasing verb	context	L1 adult		L1 child		L2 adult	
		NP1	NP2	NP1	NP2	NP1	NP2
empören (outrage)	diver/pirate	**21**	3	**19**	5	**23**	3
	salesman/garbageman	**19**	5	**13**	11	**15**	10
beeindrucken (impress)	gardener/farmer	**22**	2	**18**	6	**28**	3
	prince/king	**23**	1	**21**	2	**27**	4
*begeister*n (inspire)	construction worker/firefighter	**22**	2	**21**	3	**25**	6
	king/prince	**23**	1	**15**	7	**27**	4
beunruhigen (agitate)	cowboy/Native American	**22**	2	**21**	3	**24**	7
	bandit/sheriff	**24**	0	**20**	3	**27**	4
erschrecken (scare)	policeman/burglar	**24**	0	**23**	1	**27**	4
	guard/knight	**23**	1	**21**	3	**25**	6
verwirren (confuse)	firefighter/construction worker	**21**	3	**20**	3	**21**	9
	doctor/pilot	**23**	1	**24**	0	**25**	4
überraschen (surprise)	clown/musician	**24**	0	**22**	1	**28**	3
	garbageman/salesman	**23**	1	**22**	2	**29**	2
ängstigen (frighten)	Native American/cowboy	**22**	2	**22**	2	**25**	6
	sheriff/bandit	**21**	3	**20**	3	**27**	4
erfreuen (delight)	Roman/Egyptian	**24**	0	**19**	5	**28**	3
	pilot/doctor	**24**	0	**23**	1	**28**	3
verärgern (anger)	pirate/diver	**24**	0	**19**	4	**24**	7
	knight/guard	**24**	0	**18**	6	**25**	6
verblüffen (amaze)	burglar/policeman	**23**	1	**22**	2	**23**	1
	musician/clown	**23**	1	**19**	5	**18**	6
erstaunen (astonish)	Egyptian/Roman	**21**	3	**18**	6	**20**	11
	farmer/gardener	**20**	4	**15**	9	**14**	17

Table C.4: Total number of NP1 and NP2 decisions per group for each context with an NP1-biasing verb in Pre-test 2 for Experiment 4

NP2-biasing verb	**context**	**L1 adult**		**L1 child**		**L2 adult**	
		NP1	NP2	NP1	NP2	NP1	NP2
bedauern (be sorry for)	knight/guard	4	**20**	6	**18**	9	**21**
	sheriff/bandit	10	**14**	11	**13**	8	**21**
vergöttern (adore)	clown/musician	7	**17**	7	**17**	7	**23**
	prince/king	0	**24**	4	**20**	4	**26**
bemitleiden (pity)	cowboy/Native American	5	**19**	7	**15**	4	**25**
	farmer/gardener	6	**18**	4	**20**	9	**20**
mögen (like)	Egyptian/Roman	2	**22**	4	**19**	3	**28**
	doctor/pilot	1	**23**	3	**19**	6	**25**
fürchten (fear)	Native American/cowboy	1	**23**	5	**19**	8	**23**
	burglar/policeman	3	**21**	4	**20**	13	**18**
bewundern (admire)	gardener/farmer	5	**19**	3	**21**	7	**24**
	king/prince	3	**21**	2	**22**	3	**28**
hassen (hate)	policeman/burglar	3	**21**	6	**17**	6	**24**
	guard/knight	4	**20**	0	**24**	8	**23**
verachten (despise)	diver/pirate	4	**20**	5	**19**	5	**23**
	salesman/garbageman	3	**21**	3	**21**	7	**21**
bevorzugen (prefer)	Roman/Egyptian	2	**22**	5	**19**	12	**18**
	garbageman/salesman	4	**20**	5	**19**	5	**25**
lieben (love)	musician/clown	0	**24**	1	**23**	5	**26**
	firefighter/construction worker	3	**21**	6	**18**	6	**25**
respektieren (respect)	pirate/diver	4	**20**	4	**20**	4	**27**
	bandit/sheriff	5	**19**	3	**21**	2	**29**
verehren (worship)	construction worker/firefighter	2	**22**	5	**18**	10	**18**
	pilot/doctor	6	**18**	10	**14**	11	**17**

Table C.5: Total number of NP1 and NP2 decisions per group for each context with an NP2-biasing verb in Pre-test 2 for Experiment 4

Subject	Age	NP1 bias strength	NP2 bias strength	filler accuracy	NP1	NP2
1	9;7	50%	39%	73%	79%	67%
2	9;10	95%	71%	100%	100%	100%
3	10;2	52%	52%	54%	54%	54%
4	10;3	45%	73%	98%	96%	100%
5	8;11	68%	55%	85%	96%	75%
6	9;8	55%	45%	25%	33%	16%
8	10;10	86%	95%	96%	100%	92%
9	11;0	41%	50%	48%	46%	50%
11	12;5	82%	86%	96%	100%	92%
12	11;2	100%	91%	90%	100%	79%
13	12;5	100%	100%	96%	100%	92%
14	10;7	86%	74%	58%	71%	46%
15	10;3	90%	86%	92%	96%	88%
16	10;8	64%	55%	81%	79%	83%
17	9;6	95%	50%	90%	100%	79%
18	12;10	95%	91%	100%	100%	100%
19	12;5	77%	52%	60%	67%	54%
20	12;7	91%	19%	60%	100%	21%
21	9;9	91%	32%	69%	92%	46%
22	11;2	82%	89%	98%	96%	100%
23	12;9	86%	82%	96%	100%	92%
24	13;0	76%	86%	92%	88%	96%
25	12;4	91%	64%	100%	100%	100%
26	12;9	90%	95%	96%	100%	92%
27	10;11	95%	95%	98%	96%	100%
28	10;7	86%	86%	100%	100%	100%
29	12;2	86%	64%	94%	96%	92%
30	13;0	50%	73%	96%	100%	92%
31	12;2	100%	100%	98%	100%	96%
32	12;10	82%	82%	100%	100%	100%
33	10;10	92%	13%	56%	96%	17%
34	8;11	100%	0%	58%	100%	17%
35	11;11	82%	50%	69%	88%	50%
36	8;6	57%	86%	90%	83%	96%
37	10;4	95%	86%	100%	100%	100%

Table C.6: Results per child in Pre-test 2 for Experiment 4. Children considered as non-comprehenders based on their filler accuracy are highlighted. The filler accuracy is also shown separately for the type of reference, i.e., whether the correct referent was the NP1 or NP2. Note that Subject 7 and 10 are not listed due to prior exclusion as indicated in the participant section.

Erklärung

Gemäß §4 (2) 4. und §4 (2) 7. der Promotionsordnung der Humanwissenschaftlichen Fakultät der Universität Potsdam vom 15.05.2013.

Hiermit erkläre ich, Judith Schlenter, dass ich die vorliegende Arbeit selbstständig und ohne unzulässige Hilfe Dritter verfasst habe und bei der Abfassung alle Regelungen guter wissenschaftlicher Standards eingehalten habe.
Ich erkläre weiterhin, dass diese Dissertation in der gegenwärtigen Fassung keiner anderen Hochschule zur Begutachtung vorgelegen hat oder vorliegt, und dass ich an keiner anderen Hochschule ein Promotionsverfahren eröffnet habe.

Berlin, den 12.08.2019