Processing of Non-canonical Word Orders in an L2
-
When small changes make no big difference

by Silke Schunack, M.A.

submitted to the Faculty of Human Sciences
at the University of Potsdam
in 2016

“It is a capital mistake to theorize before one has data.
Insensibly one begins to twist facts to suit theories,
instead of theories to suit facts.”

(Sherlock Holmes – A Scandal in Bohemia, 1891)

Abstract

This thesis investigates the processing of non-canonical word orders and whether non-canonical orders involving object topicalizations, midfield scrambling and particle verbs are treated the same by native (L1) and non-native (L2) speakers. The two languages investigated are Norwegian and German.

32 L1 Norwegian and 32 L1 German advanced learners of Norwegian were tested in two experiments on object topicalization in Norwegian. The results from the online self-paced reading task and the offline agent identification task show that both groups are able to identify the non-canonical word order and show a facilitatory effect of animate subjects in their reanalysis. Similarly high error rates in the agent identification task suggest that globally unambiguous object topicalizations are a challenging structure for L1 and L2 speakers alike.

The same participants were also tested in two experiments on particle placement in Norwegian, again using a self-paced reading task, this time combined with an acceptability rating task. In the acceptability rating L1 and L2 speakers show the same preference for the verb-adjacent placement of the particle over the non-adjacent placement after the direct object. However, this preference for adjacency is only found in the L1 group during online processing, whereas the L2 group shows no preference for either order.

Another set of experiments tested 33 L1 German and 39 L1 Slavic advanced learners of German on object scrambling in ditransitive sentences in German. Non-native speakers accept both object orders and show neither a preference for either order nor a processing advantage for the canonical order. The L1 group, in contrast, shows a small, but significant preference for the canonical dative-first order in the judgment and the reading task.

The same participants were also tested in two experiments on the application of the split rule in German particle verbs. Advanced L2 speakers of German are able to identify particle verbs and can apply the split rule in V2 contexts in an acceptability judgment task in the same way as L1 speakers. However, unlike the L1 group, the L2 group is not sensitive to the grammaticality manipulation during online processing. They seem to be sensitive to the additional lexical information provided by the particle, but are unable to relate

the split particle to the preceding verb and recognize the ungrammaticality in non-V2 contexts.

Taken together, my findings suggest that non-canonical word orders are not per se more difficult to identify for L2 speakers than L1 speakers and can trigger the same reanalysis processes as in L1 speakers. I argue that L2 speakers' ability to identify a non-canonical word order depends on how the non-canonicity is signaled (case marking vs. surface word order), on the constituents involved (identical vs. different word types), and on the impact of the word order change on sentence meaning. Non-canonical word orders that are signaled by morphological case marking and cause no change to the sentence's content are hard to detect for L2 speakers.

Keywords: L2 sentence processing, object topicalization, scrambling, particle verbs, Norwegian, German, self-paced reading, acceptability judgments

Acknowledgments

Four years of work have finally come to an end and have culminated in this thesis that consumed my passion for foreign languages and science, and apparently also my youth as I recently discovered plenty of grey hairs that were not present at the start of this endeavor. What had seemed like the right thing, actually the *one* thing, to do for 20 years from elementary school to pretty much the first year of my PhD turned into something that can best be described by a quote from Heinrich Heine: “Anfangs wollt ich fast verzagen, und ich glaubt, ich trüg es nie; und ich hab es doch getragen - aber fragt mich nur nicht, wie?” The non-German speaking reader is referred to machine translation programs for a native language adaptation, I’m sure it will be hilarious. There have been many people who contributed to the unique experience of this PhD and they will be acknowledged here.

First, my supervisors: Claudia Felser, for accepting me as her PhD student, for her extensive comments on complete and pre-final drafts of this thesis, and for sticking with me in the past four years. Marit Westergaard, for being my second supervisor and helping me with the Norwegian section of this thesis, and the data acquisition in Tromsø. Michaela Schmitz, my unofficial third supervisor, who was supportive and knowledgeable to no end whether in the quarterly supervisory board meetings or in matters of funding for traveling or participants.

Second, the funding parties: Participant and traveling funds for this thesis were provided by the Potsdam Graduate School, the Kommission für Forschung und wissenschaftlichen Nachwuchs (FNK) of the University of Potsdam and the Potsdam Research Institute for Multilingualism (PRIM). During the writing phase I was partly supported by a grant by PRIM. Day-to-day expenses were secured by a one-year-contract as Wissenschaftliche Hilfskraft at PRIM which allowed me to gain insight into research methods beyond the ones used in this thesis. And by a seasonal contract with the Filmpark Babelsberg which helped me experience the world outside the ivory tower of science and made me explain my work to non-scientists many, many times. The most important funding party, however, is my mother. Without her far-sightedness regarding saving money for my education since the day I was born, the past four years would not have been possible or would have taken eight.

Third, the people. On the scientific side, I have to mention my colleagues at PRIM. All you PhDs, post docs and professors that were part of the team since I joined. I guess I profited from every one of you in different ways. Be it comments on my presentations in the lab meetings or help with experimental setups and data analysis, or sharing the stories about success and failure that we all experienced. Parts of this thesis have been presented at conferences and I am thankful for the comments on my work and the chats over lunch.

This thesis was inspired by the many teachers of foreign languages who opened my eyes to the multitude of languages in this world and the fascinating cultures expressed by them. Even if foreign language learners might not be like native speakers from a processing perspective, they can learn to love any language like their own as I witnessed in many of my wonderful participants.

Personal thanks to Marius for his kind help with the Norwegian materials and to Sara Ann for her lightning fast proofreading of the final version of this thesis.

On the social side, I could rely on the support of a number of people. My coaches from the university's sports program, especially Kata, Svenja, Sarah and Sebastian, for not only taking care that my body stays healthy over the length of this project (almost no back pain despite many hours of sitting on desks and staring at a computer screen!), but also ensuring a healthy mind in said body. Many scientific problems find a sudden solution after 60 minutes of zumba, 90 minutes of boxing exercise or after carrying a grown-up person on your back around a track. I would assume that it is the additional oxygen that hits your brain and the fact that when you reach or even surpass your physical limits p-values do not seem that important anymore.

Big thanks to my unofficial PhD support group: Martín, Haris, Yan, and especially in the later phases, Carla and Antje. You come from different fields (IT, medicine, engineering and linguistics) and you completed, quit, or are still writing your PhD (the reader may guess who is who). You have been incredible at making me smile again when I needed it, at putting things back in perspective when I couldn't do it anymore and most of all – you did not let me quit. If it hadn't been for your social supervision of my PhD, I probably would have quit long ago.

To David, you have been more supportive than you can imagine, because I could dump all my problems and insecurities on you, write pages of complaints only to learn that sometimes the answer to a long, dragged-out question can be as simple as Yes., No. or ☺. The few weeks I could spend with you each year were always precious and filled me with new energy to complete this massive task. You pushed me over my personal boundaries time and time again and I learned how much stress I can take physically and mentally without giving in. Thank you for reaching out your hand to me 5.5 years ago, and promising me you would not let me fall - you have kept this promise until today.

Finally, the most important thank you goes to my mother. Without your unconditional support none of this would have been possible. You made sure that I could try out whatever talents I thought I had and you always had my back and supported any of my decisions, even the one for linguistics despite the presence of other options with better job security and probably better payment. I was the first woman in my family to attend university and I am the first family member to receive a PhD. Needless to say, I am proud of this achievement.

Table of Contents

List of Abbreviations

A/ACC – accusative

Appl – applicative

AoA – age of acquisition

BNF – beneficiary

CM – Competition Model

D/DAT – dative

DO – direct object

DOC – Double Object Construction

DP – determiner phrase

EIC – Early Immediate Constituents

ERPs – event-related potentials

FIH – Fundamental Identity Hypothesis

fMRI – functional magnetic resonance imaging

INST – instrumental

IO – indirect object

L1/L2 – first/second language, native/non-native language

LAN – left anterior negativity

LOC - locative

NP – noun phrase

NOM – nominative

O – object

OCTT – Oslo Corpus of Tagged Texts

P – phrase

PAST - past tense

PDC – Prepositional Dative Construction

PP – prepositional phrase

PRES – present tense

PRET - preterite

REC – recipient

SFA – syntactic function ambiguity

SLA – second language acquisition

SPEC – specifier

SPR – self-paced reading

SSH – Shallow Structure Hypothesis

S/SUBJ – subject

t - trace

V – verb

VP – verbal phrase

List of Figures

List of Tables

1 Introduction

“Named must your fear be before banish it you can.”
(Yoda, Star Wars Episode V: The Empire Strikes Back, 1980)

The above quote from a popular character of an even more popular movie franchise is an illustration of the core topic of this dissertation – word order variation. Altering the linear order of words in a sentence can have many purposes such as highlighting or backgrounding information, making a sentence’s intonation smoother, or a humorous or poetic effect. The range of possible word orders is language-specific and any grammatical order can be derived from that language’s grammar. Analyses by Liberman (2005) and Pullum (2005) find that even the ungrammatical word order that the character Yoda uses follows certain ordering rules and is not a random arrangement of words.

Word orders can also be permutated to different degrees. Slight changes often go along with only small changes to meaning as in the two possible alternations for English particle verbs in (1a) and (1b). Both sentences can be easily understood as describing an action that involves a male person, a new computer and the process of starting that computer (see Haddican & Johnson, 2012 for a comparison of ordering preferences in British and American English). Compare this to contrastive word orders that topicalize one element such as (1c) and (1d).

(1a) David turns on his new computer.

(1b) David turns his new computer on.

(1c) On did David turn his new computer.

(1d) His new computer did David turn on.

While still describing the exact same action, (1c) involves contrasting the action of turning the computer on with the action of turning it off and (1d) contrasts the new computer with some other entity, for example the TV. These two sentences are already harder to understand and higher in syntactic complexity than (1a) and

(1b). The complexity could in theory be further increased to Yoda-like levels, and languages other than English indeed allow a lot more word order flexibility that creates sentences that are highly challenging to the parser.

Why use complex sentences with unusual, non-canonical word orders that are hard to understand, if there is always the possibility to use the easier canonical word order? Why do these challenging orders persist in language? A study by Kristensen, Engberg-Pedersen, & Poulsen (2014) on object topicalization in Danish tries to give an answer. While non-canonical word orders by themselves are harder to understand and parse than canonical ones, it is their special ability to draw the hearer's attention towards a specific element that makes them valuable in written and spoken language. Unlike in experimental settings, non-canonical sentences usually do not appear without a supporting context that justifies the non-canonicity. This context helps to figure out the meaning of non-canonical word orders. So, it is the context and the special information status that keep non-canonical word orders alive.

This thesis presents four studies on non-canonical word orders. It focuses on the processing of said word orders and not on the influence of context. Most aspects of this thesis can be divided into two factors:

Two languages were investigated: Norwegian and German. Research on German canonical and non-canonical word orders is comparatively common and has produced a large body of experimental findings (Bader & Meng, 1999; Gerth, Otto, Felser, & Nam, 2015; Hopp, 2005; Jackson, 2007; Pappert, Schließer, Janssen, & Pechmann, 2007; Rösler et al., 1998; Schlesewsky, Bornkessel, & Frisch, 2003). The language lends itself nicely to a number of order manipulations due to its case marking system which allows constituent orders beyond that of English, the most used language in psycholinguistic research, which has a stricter word order and does not allow the same order manipulations as German. This syntactic difference regarding the investigated order manipulations complicates the comparison between English and German. Research on Norwegian has so far been mainly limited to theoretical approaches (Askedal, 1984; Bentzen, 2007a, 2007b; Fretheim, 1992) and first language acquisition studies (Anderssen, Bentzen, Rodina, & Westergaard, 2010). Processing studies on any of the related Scandinavian languages (Danish, Swedish or Icelandic) are also extremely rare (Kristensen, Engberg-Pedersen, Højlund Nielsen, & Wallentin, 2013; Kristensen et al., 2014; Roll, Horne, & Lindgren, 2007). Norwegian is, however, a very interesting language to study with regards to word order variation. Like English, it lacks case marking on non-pronominal noun phrases and verbs are only marked for tense. Despite this limited morphology, it does allow some word order operations that are not possible in English. This way Norwegian fits neatly in a gap between English and German, because in some ways it can behave like German with a morphological make-up that is similar to English.

Two groups of participants took part in the experiments: native (L1) and non-native (L2) speakers.[1] The field of native language processing has been fruitfully studied for an extended time now and countless theories of native language processing have been developed based on this research. Non-native language processing research has seen a sharp rise in the last 25 years with more and more researchers and studies turning towards this population to enrich their knowledge. Non-native speakers may struggle with constructions that pose no problems to native speakers. Their processing difficulties can help linguists to improve their models of language processing and comprehension. This thesis adds to the growing body of L2 research by focusing on questions of L2 syntactic processing in Scandinavian languages that have previously received attention mainly in questions of phonology and vocabulary (Jin, Åfarli, & van Dommelen, 2007; Rydland, Grøver, & Lawrence, 2013; van Dommelen & Husby, 2007), but see Abrahamsson & Hyltenstam (2008) and Abrahamsson (2012) for extensive research on various aspects of near-native L2 Swedish. At the same time, due to the scarcity of studies investigating native Norwegian sentence processing, collecting native speaker data contributes entirely new results for this language. As the body of research on L2 German is larger than for L2 Norwegian the contribution of my thesis in this area is less of the pioneering kind and rather adds to existing findings by investigating non-native speakers from an L1 background that has previously not been considered for the phenomena in this thesis.

[1] In this thesis, I use the term L2 to refer to any foreign language acquired after the native language. It does not imply that this was the second language in order of acquisition. The biographical data provided in Appendix C lists the actual chronological number of the L2 invesitigated under LX.

Two main methods have been used to complement each other. The main psycholinguistic experimental tool employed in this thesis was a self-paced reading (SPR) paradigm to assess the participants' online sentence processing via the measurement of their reading times. Questionnaires with acceptability judgment tasks and, in one case, an agent identification task were used to assess the untimed comprehension of sentences and participants' general sensitivity to the experimental manipulations. The combination of these two methods is especially interesting and almost necessary for the L2 group as previous studies have found different results in online and offline tasks for non-native speakers (e.g. Coughlin & Tremblay, 2012; Patterson, Trompelt, & Felser, 2014). These results often showed that knowledge about the manipulation, e.g. an ungrammaticality, is present in the untimed, offline task, but absent in the timed, online task in which the participants are under pressure to perform quickly. The offline questionnaires also aimed to improve our understanding of what is considered acceptable by native speakers. Traditionally, assumptions about grammaticality and acceptability have been based on judgments made by theoretical linguists that were published in theoretical papers or descriptive grammars. With the advent of tools that allow the collection of judgments of a large number of speakers with comparatively little effort, some researchers in generative syntax research suggest that grammaticality and acceptability judgments should also be collected in a rigorous experimental manner (see Gibson & Fedorenko, 2013 for a discussion). The judgments collected in this thesis follow this suggestion in order to experimentally verify claims by descriptive grammar, especially for Norwegian.

Two main phenomena build the core of this thesis. One of these is non-canonical object position, which can be further subdivided into object topicalization, in which objects are found in a sentence-initial position (2a,b), and scrambling, in which objects change their order in the German midfield (3a,b). Topicalization and scrambling have been studied in a range of languages and using a number of techniques, including eye tracking and ERPs.

(2a) Mannen vil kysse kvinnen. (canonical order)
man.the want$_{\text{PRES}}$ kiss woman.the
'The man wants to kiss the woman.'

(2b) Kvinnen vil mannen kysse. (topicalized object)
woman.the want$_{\text{PRES}}$ man.the kiss
'The woman, the man wants to kiss.

(3a) Die Enkelin hat dem Großvater den Kuchen mitgebracht.
(canonical order)
the$_{\text{NOM}}$ granddaughter has the$_{\text{DAT}}$ grandfather the$_{\text{ACC}}$ cake brought
'The granddaughter has brought the grandfather the cake.'

(3b) Die Enkelin hat den Kuchen dem Großvater mitgebracht.
(scrambled accusative)
the$_{\text{NOM}}$ granddaughter has the$_{\text{ACC}}$ cake the$_{\text{DAT}}$ grandfather brought
'The granddaughter has brought the cake to the grandfather.'

The second phenomenon is particle verbs that have been widely discussed in theoretical works and with regard to their lexical storage, but less so in experimental studies. The Norwegian particle verb structures allow the same variation as the English examples (1a) and (1b) and could also be considered as being instances of objects in non-canonical positions depending on which structure is assumed to be the basic one. German particle verbs function differently and show an interaction of verb morphology and syntax in that the particle is either split from the verb (4a) or attached as a prefix (4b), depending on the syntactic context.

V2 context, split particle
(4a) Martin isst den Hamburger auf.
Martin eats the$_{\text{ACC}}$ hamburger up
'Martin eats up the hamburger.'

non-V2 context, prefixed particle
(4b) Martin will den Hamburger aufessen.
Martin wants the$_{\text{ACC}}$ hamburger up.eat
'Martin wants to eat up the hamburger.'

The two phenomena are not equally complex across the two languages regarding their syntactic structure and their impact on sentence interpretation. This difference causes varying degrees of processing difficulty across the four studies. The manipulation used in the studies on German scrambling and Norwegian particle verbs creates only a subtle difference between the two orders without causing a change to the content or to the grammaticality. The manipulation in the Norwegian topicalization study also produces only grammatical sentences, but bears the inherent risk of misinterpreting the object-first sentence as a subject-first sentence and thereby comprehending the exact opposite of the actual content. The study on German particle verbs on the other hand manipulates the grammaticality of the sentence, while not changing its content. With this design, it will be possible to compare more obvious manipulations (grammaticality, content) with less obvious word order changes (preferences) and whether they are treated in the same way by native and non-native speakers.

Based on this kind of 2x2x2x2 design this thesis contains four studies (two languages, two phenomena) with altogether eight experiments (two methods) and sixteen datasets (two populations) that were analyzed. Table 1.1 summarizes the design and relates it to the chapters in the thesis.

		Norwegian		German	
		L1	L2	L1	L2
Non-canonical objects	Judgment task	Experiment 1a (Chapter 4.3)		Experiment 2a (Chapter 5.2)	
	SPR	Experiment 1b (Chapter 4.4)		Experiment 2b (Chapter 5.3)	
Particle verbs	Judgment task	Experiment 3a (Chapter 7.3)		Experiment 4a (Chapter 8.2)	
	SPR	Experiment 3b (Chapter 7.4)		Experiment 4b (Chapter 8.3)	

Table 1.1 Overview of experimental design and corresponding dissertation chapters

The following overarching research questions will be addressed:

Q1 Are non-canonical word orders generally more difficult to process than canonical ones?

a) Does this processing difficulty surface for all types of word order variation investigated (topicalizations, scrambling, particle verbs)?

b) Is there a difference between native (L1) and non-native (L2) processing of word order variations?

c) Does the non-native parser detect all order manipulations equally well? Does it use the same cues as the native parser to identify non-canonical orders?

Q2 Are non-canonical word orders perceived and rated as less acceptable than canonical ones?

a) Is the difference in acceptability the same for all types of word order variation investigated or are some more comparable in acceptability than others?

b) Can native and non-native speakers identify gradient acceptability equally well or is a bimodal distinction of grammatical/ungrammatical easier to identify?

Q3 Are patterns or preferences that emerge in the offline tasks also reflected in the online processing behavior?

a) Does native-like performance of the L2 group in the offline task equal native-like performance in online processing?

b) Does the ability to achieve a native-like pattern depend on the type of word order variation?

These are intentionally very broad research questions that can be applied to all structures investigated and to all tasks used. More refined research questions can be found in the introductory sections of the respective experimental chapters.

This dissertation is organized in the following way. Chapter 2 provides the reader with general background information. The first two sections review some of the current sentence processing models in L1 research (Section 2.1) and in L2 research (Section 2.2). This is followed by Section 2.3 that discusses theoretical approaches to word order phenomena and mainly contrasts derivational and non-derivational models. Section 2.4 aims to familiarize the reader with the experimental methods used. As two different phenomena were investigated in this thesis, it is divided into two bigger parts – Part I and Part II. Part I contains Chapters 3 to 5 and is concerned with the topic of objects in non-canonical positions. Chapter 3 introduces the reader to the common background that topicalization and scrambling have and to less experimental studies such as corpus studies (Section 3.1). Section 3.2 reviews the experimental literature with regard to subject/object ambiguities and garden path studies (Section 3.2.1), studies on word order in German embedded sentences (Section 3.2.2) and studies investigating the application of linearization patterns in ditransitive sentences (Section 3.2.3). Chapter 4 presents Study 1 on object topicalization in Norwegian. It begins again with a short background chapter (Section 4.1), followed by the results of a pilot study (Section 4.2). Sections 4.3 and 4.4 report Experiment 1a, the offline agent identification task, and Experiment 1b, the online SPR task. The conclusion in Section 4.5 seeks to address the research questions based on the results of the experiments. Chapter 5 contains Study 2 on object order in German ditransitive sentences. Section 5.1 is an extensive introduction that describes the German background and the ordering preferences in the Slavic native languages of the L2 group. Sections 5.2 and 5.3 report Experiment 2a, the acceptability judgment task, and Experiment 2b, the SPR task. Chapter 5 closes with a discussion of the results of Study 2 (Section 5.4) and an intermediate conclusion that addresses the results of Studies 1 and 2 and closes Part I of this dissertation (Section 5.5). Part II contains Chapters 6 to 8. Chapter 6 is concerned with the theoretical background of particle verbs (Section 6.1) and reviews the literature on Norwegian object and particle shift (Section 6.2.1), and the processing literature on non-local dependencies of which particle verbs are one example (Section 6.2.2). Chapter 7 contains Study 3 on Norwegian particle verbs. Section 7.1 again provides the theoretical background. The results of a pilot study can be found in Section 7.2. The acceptability rating task of Experiment 3a is reported in Section 7.3 and the corresponding SPR task in Section 7.4 that is followed by a conclusion (Section 7.5). Chapter 8 contains the fourth study on German particle verbs. The background information is contained in Section 8.1. Sections 8.2 and 8.3 report Experiment 4a (acceptability judgment) and Experiment 4b (SPR task). Two conclusions, one specific to Study 4 (Section 8.4), the other an intermediate conclusion (Section 8.5), comprising the results of both studies on particle verbs, round up Chapter 8 and Part II. The general discussion in Chapter 9 brings together the results of all four studies and discusses their findings with relation to the processing models reported in Sections 2.1 and 2.2 and the general research questions presented above. It also contains suggestions for future research based on new questions raised by the results of the reported experiments (Section 9.1) and contains the conclusion in Section 9.2. In addition to these content chapters there are three appendices. Appendix A contains all the materials used in the experiments and the pilot studies, including fillers. Additional background data, alternative analyses, and more in-depth renditions of the results reported can be found in Appendix B. Appendix C gives an anonymized overview of the participants in the four studies.

The main findings of the four studies reported are in line with research and theories assuming a difference in processing behavior between native and non-native speakers even at high proficiency levels. In three out of four online studies L2 participants showed a processing pattern that deviated from the one of the L1 group. However, they showed a nativelike processing pattern in the study on object topicalizations that required a thorough syntactic and thematic reanalysis and contained a syntactically complex non-local dependency. L2 participants were more successful in identifying non-canonical structures in the offline tasks. They showed the same elevated error rates for object topicalizations in the agent identification task as the L1 group. They also exhibited similar general ordering preferences as the L1 group in the acceptability rating task. However, the L2 group did not show evidence for a gradient acceptability found in the acceptability ratings of the L1 group. Taken together, these findings suggest that failure to identify and process a non-canonical sentence is not directly related to a sentence's syntactic complexity. Instead, the detection of the non-canonicity seems to be easier for L2 speakers, if it can be identified using a shallow parse without morphological specifications, as is the case for Norwegian object topicalizations or Norwegian particle verbs, and if the non-canonical order requires a reanalysis of the content of the sentence. Non-canonical orders that are signaled by morphological case marking on objects and are reflections of a small change in information structure are harder to detect for L2 speakers and are not reflected in different online processing patterns.

2 General background

This chapter will provide the reader with the general background knowledge needed to grasp the linguistic concepts and theories explored in this thesis. It is intended to give an overview, rather than a detailed discussion, of competing processing models or theories. For more in-depth discussions of the syntactic phenomena investigated in this thesis, the reader is referred to the respective experimental chapters. This chapter consists of four sections. Section 2.1 introduces and contrasts the most popular models of sentence processing, ranging from models focused mainly on syntax to memory-based models. Section 2.2 looks at hypotheses on second language (L2) processing that center on the discussion whether L1 and L2 processing are identical. As this thesis has non-canonical object placement as one of its topics, Section 2.3 gives a short description of movement-based approaches to language representation compared to base-generation approaches. Section 2.4 covers the experimental methods used in this thesis. It is aimed at readers that are less familiar with psycholinguistic techniques and discusses the rationales behind each method, including its strengths and its weaknesses. Section 2.5 concludes this section with a short summary.

2.1 Introduction to processing models

Theories of sentence processing are concerned with the interaction and integration of various sources of information (i.e. syntax, semantics, pragmatics, discourse and context) during online language performance in comprehension as well as in production. One main theoretical divide concerns the role syntax plays in processing compared to the other available sources of information: **garden path or syntax-first models** (e.g. Frazier & Fodor, 1978) highlight the use of syntactic information independent of semantic or pragmatic information, while **interactionist and constraint-based models** (e.g. MacDonald, 1997; McClelland, John, & Taraban, 1989) assume the immediate and interactive integration of all available sources of information. These two different approaches will be discussed in Sections 2.1.1 and 2.1.2 respectively.

Another area of debate concerns the nature of the parser. One side assumes a parser that is universal to all languages and all speakers (e.g. Frazier & Fodor, 1978), whereas the other side assumes an exposure-based parser that reflects the experience of the speaker with a particular language, and draws heavily on the frequency of occurrence of a particular structure to explain its acquisition and processing (Mitchell, Cuetos, Corley, & Brysbaert, 1995). As exposure-based parsers usually assume an exposure to different kinds of information, **experience-based models** are also included in Section 2.1.2. **Memory-based processing** accounts that draw upon general cognitive principles and **'good enough' processing models** that assume incomplete parsing are briefly touched upon in Sections 2.1.3 and 2.1.4.

2.1.1 Garden path or syntax-first models

Garden path models assume that syntactic analyses are constructed in a serial fashion, i.e. only one interpretation is sustained at a time and all other possible analyses are discarded. As the second name for these types of models ('syntax-first') suggests, syntax is the main source of input in these models and additional information from other sources such as context or plausibility is not considered in the first parsing attempt. However, this additional information is not completely disregarded, it is used to evaluate the chosen interpretation once the parser has settled for it. According to these types of models, the parsing process is governed by general principles and heuristics that aim to minimize the effort put into processing and to decide between several structures that are equally licensed by the grammar.

The principle of **Minimal Attachment** encourages the building of the simplest structure possible and the avoidance of revisions. For the SVO and SOV languages investigated in this thesis this means the following: When the parser for example encounters a sentence-initial NP that could either be interpreted as a subject or an object, it will interpret this NP as a subject, because in transformational grammar a subject-initial sentence has a simpler syntactic structure than an object-initial sentence.[2] **Late Closure**, another principle,

[2] A non-transformational perspective on sentence structure would possibly also predict a preference for the subject-initial sentence instead of an object-initial sentence, but not based on its lower syntactic complexity. As non-transformational grammars do not derive surface structures

postulates the incorporation of incoming material into the current processing domain instead of the retrieval of old constituents from memory or the construction of a new constituent. If the parser reaches for example a word like *on* that could either be a particle and be integrated into the current verbal phrase (VP, 5a) or a preposition that is part of a new prepositional phrase (PP, 5b), Late Closure would predict an interpretation as a particle.

(5) Yan puts the hat on….
(5a) … because it is cold. (particle interpretation)
(5b) … the table. (prepositional interpretation)

The aim to keep the distance between two connected constituents (e.g. a filler and its gap) as short as possible is contained in the **Minimal Chain** principle or the **Active Filler Hypothesis**. The parser will try to posit a gap and reactivate the filler as early as possible in a sentence. In example (6) below (taken from Williams, Möbius, & Kim, 2001) this means that a first gap will be assumed after the word *push* making the wh-element the object of the verb. This analysis has to be abandoned subsequently as the gap is filled by *the bike* as the actual object of the verb and the correct gap is part of the following PP. Filled-gap effects have been investigated using stop-making-sense tasks or cross-modal priming that investigates the reactivation of the filler at possible gap sites.

(6) Which girl did the man push ________ the bike into ______ late last night?
filler false gap gap

Two further principles govern the revision process. **Revision-as-last-resort** states that revisions are only undertaken by the parser if the previously developed syntactic representation is wrong in order to change it. The principle of **Minimal Revision** states that repairs should use the fewest changes to fix the representation and not attempt to create an ideal representation.

This group of models takes its names from the **garden path effect** that is observed when the parser computes a preferred syntactic analysis that turns out to be incorrect when disambiguating material is encountered. The initial analysis then needs to be discarded or revised (‘reanalyzed’) in favor of the correct,

from a deeper syntactic structure and instead generate them directly, the principle of Minimal Attachment would be largely limited to the avoidance of revisions.

initially dispreferred syntactic interpretation. Example (7) below illustrates an ambiguity involving a reduced relative clause and example (8) illustrates a direct object ambiguity.

(7) The horse raced past the barn fell.

(8) Peter knew the answer was false.

In example (7), the parser initially interprets the sentence as a regular main clause with *the horse* as the subject, *raced* as the main verb and *past the barn* as a prepositional phrase. When encountering *fell*, the parser has to reinterpret *raced past the barn* as a reduced relative clause attached to the subject, and *fell* is then interpreted as the new main verb of an intransitive clause. In example (8), *the answer* is at first interpreted as a direct object of the transitive verb *knew*. When the disambiguating material *was* is encountered, *knew* has to be reinterpreted as intransitive and the interpretation of *the answer* has to be changed from object of a main clause to subject of an embedded clause. The point at which the disambiguating information occurs, triggering reanalysis, is usually used for measuring garden path effects in psycholinguistic experiments. Garden path effects have been found to vary in strength suggesting that reanalysis is easier in some contexts than in others and leading to a division of weak and strong garden paths. Both types trigger a reanalysis and cause elevated reading times in SPR tasks and backtracking in eye tracking studies, but only strong garden paths also affect accuracy on comprehension questions and are often judged as ungrammatical.

A second definition of garden path sentences that is more wide-spread in non-serial models of sentence processing only entails sentences that cause conscious processing difficulties (Weinberg, 1993). Example (8) would not be a garden path in this definition as it does not cause conscious processing difficulties. For this thesis I adopt the first, more general definition as it is not possible to differentiate between conscious and unconscious processing difficulty with the experimental methods that I used.

The process of reanalysis clearly involves the revision of a previous analysis, but what triggers this revision and what exactly is revised varies between different authors. Strict accounts of garden path models assume a purely syntax-driven form of reanalysis that affects and changes the syntactic representation of the sentence, but there have been suggestions for additional types of reanalysis. Bornkessel, Schlesewsky, & Friederici (2003) suggest an extension of the definition of reanalysis to also include the possibility of thematic reanalysis. Their definition of reanalysis is the following: "(…) the need to perform some sort of a *recomputation* with regard to the analysis of a linguistic input that has been built up so far." (Bornkessel et al. 2003:270, italics in original). The experimental evidence supporting this definition will be reviewed in Section 3.2.

Bader (2000) proposes that the ease of recovery from a syntactic misanalysis does not solely depend on the syntactic information received, but also on lexical-morphological and prosodic information. Assuming a modular architecture of the human sentence processor, the phonological coding module aggravates the reanalysis process whenever the prosodic structure of a sentence needs to be revised in addition to a revision of the syntactic structure (Prosodic Constraint on Reanalysis, Bader 2000:201). The phonological module only plays a role in reanalysis for reading comprehension as the reader has to build a phonological representation himself. In tasks involving auditory input, an intonational pattern is forced upon the participant, which can either match the participant's expected intonation or not. If a sentence in an auditory task begins with an NP that has contrastive stress, the listener is more likely to immediately interpret it as a sentence containing a topicalization. In a reading task, the reader decides which intonation to apply, and in the case of a topicalization without biasing context, likely has to revise the neutral intonation assigned to subject-first sentences.

The lexical module is involved in the processing of syntactic ambiguities that are caused by lexical ambiguities, e.g. through ambiguous case marking. In (9) the relative pronoun *die* and the NP *das Kind* are ambiguous between nominative and accusative case leaving the whole sentence ambiguous between a subject relative clause and an object relative clause. This only applies to a subset of sentences in German that contain words from case paradigms with high syncretism (especially feminine and neuter gender). In Norwegian global ambiguities of this kind are more widespread. Norwegian simple tense sentences with two NPs as in (10) are always ambiguous between a subject-first and an object-first interpretation due to the complete absence of case marking. Only stress and context can differentiate the two readings.

(9) Das ist die Frau, die$_i$ (t$_i$) das Kind (t$_i$) besucht hat.
this is the$_{\mathrm{NOM/ACC}}$ woman who$_{\mathrm{NOM/ACC}}$ the$_{\mathrm{NOM/ACC}}$ child visited has
'This is the woman who has visited the child/who the child has visited.'

(10) Mannen kysser kvinnen.
man.the kiss$_{\mathrm{PRES}}$ woman.the
'The man kisses the woman./The man, the woman kisses.'

Bader also states that reanalysis should not be the only option available to the parser in serial processing models. Instead, the option to not attempt reanalysis at all and reject a sentence as ungrammatical should be included in order to better account for differences in garden path strength. Experimental evidence in support of this modified definition of reanalysis will be reviewed in Section 3.2.
Strict accounts of syntax-first models have been questioned and criticized by studies that found effects of frequency, context or semantics on sentence processing. This criticism has led some researchers to admit the possibility of weak interactions between syntactic information and semantic and pragmatic information, however, the concept of syntactic autonomy has been maintained (Altmann & Steedman, 1988; Bornkessel & Schlesewsky, 2006b; Frazier & Clifton, 1997).

The predictions that syntax-first models of either variety make for the processing of canonical and non-canonical word orders are fairly straightforward. Structures that do not require a reanalysis are processed faster and more effortlessly than structures that involve a reanalysis of any kind. Non-canonical word orders often require a reanalysis of the initially specified canonical word order that should elicit elevated reading times. A prerequisite for this prediction is the correct syntactic representation of the canonical and the non-canonical order. If there is no syntactic distinction between the two orders that motivates an effortful reanalysis, e.g. in the case of a flat or under-specified representation, no processing differences would be predicted.

2.1.2 Non-syntax-centric models

Apart from syntax there are many other components of language that are featured in different types of non-syntax-centric models. In **lexicalist models** of sentence processing, the main driving force in processing is the lexicon. Detailed lexical information, such as conditional probabilities of occurrence for a lexical item, is considered if the grammar allows more than one structure. NP animacy, pronominality and concreteness are lexico-semantic properties that have all been found to influence processing. The processing of a lexical item with a dispreferred value, e.g. an inanimate subject, is more effortful than the processing of a lexical item with a preferred value, e.g. an animate subject. The processing difficulty associated with object extractions for example depends on the animacy of subject and object (Traxler, Morris, & Seely, 2002). Sentences (11a) and (11b) are both object relative clauses, but (11a) contains an animate sentential subject (*director*) and an inanimate entity in the relative clause (*movie*). (11b) on the other hand contains an inanimate sentential subject (*movie*) and an animate entity in the relative clause (*director*). The study by Traxler et al. (2002) found that object relative clauses were generally harder to process than the corresponding subject relative clauses, but the processing difficult was greatly reduced for object relative clauses with an inanimate sentential subject like (11b).

(11a) The director that the movie pleased received a prize at the film festival.

(11b) The movie that the director watched received a prize at the film festival.

The **Lexicalist Constraint-Based Model** by MacDonald (1997) and the **Competition Model** by MacWhinney and colleagues (e.g. MacWhinney & Bates, 1989) are prominent models of the lexicalist approach. They also exhibit features of **constraint-based or interactionist models** which assume a distribution of language knowledge in associated patterns and integrate all available information, be it syntax, context or frequency, into their parse. Parallel processing is often assumed in which several analyses are computed at the same time. The analysis with the most support by all information and the highest activation is then foregrounded and the other possible analyses are retained in the background to be activated in case of information that contradicts the strongest analysis. Approaches assuming parallel parsing have been criticized on the fact that the maintenance of several possible parses and the immediate integration of a wide range of available sources of information are costly and require a lot of mental effort. Populations with limited mental resources like children or L2 speakers would be at a disadvantage as they cannot retain the same number of parallel parses as speakers with more mental resources. In this section, I will present the Competition Model (CM) in more depth, since Study 1 with its manipulation of NP animacy and word order was developed from similar experiments within the CM framework. It will become evident in this section that the CM is less suitable to describe and discuss the remaining three studies.

The Competition Model was developed by MacWhinney and colleagues from the 1980s on, first for child L1 acquisition and was later adapted to bilingualism and SLA (e.g. MacWhinney, Bates, & Kliegl, 1984; MacWhinney, 1997; MacWhinney, 2005; for a general overview see Year, 2003). The model aims to account for both language acquisition and sentence processing, drawing from principles of lexical functionalist theory that emphasize the role of the lexicon in processing and as well as from connectionist modeling. This model therefore especially stresses the role of frequency in its theory of language acquisition. For the purpose of this thesis, the model's specifications for sentence processing are the most important.

The CM assumes two levels of information structure: an internal, functional level containing meanings and communicative intentions, and an external, formal level containing surface features. Language acquisition and processing involve the interaction of these two-level information structures. For lexical items, the internal function consists of their semantic properties and concepts, while the external form consists of their auditory and orthographic properties. Functional relations of lexical items are expressed through surface forms like morphological markings or word order patterns. In the CM, grammar is understood as the mapping of relational functions such as 'agent' and those surface forms. Within the model, surface forms are called 'cues' and entail "any piece of information used by listeners and speakers to determine the relationship between form and meaning" (Year 2003:7). The CM further assumes a direct relation between form and function. However, this relation is usually not a one-to-one relation, but rather a many-to-many relation, as single forms can map onto several functions and vice versa. The external form of the suffix *–s* for example can map onto the function of 'plural noun' or '3rd person singular'. In order to minimize the strain on memory resources, related forms and functions are grouped together. Some functions like topic, agent, actor and animacy are prototypically associated, as can be forms like the preverbal and the initial position in the sentence, subject-verb agreement and default stress. The mentioned forms and functions form a coalition in the network *subject,* showing that a subject consists of many associated forms and functions, and is not a single symbol. On the sentence level, the CM differentiates four different types of cues: lexical-semantic items (e.g. animacy), word order, morphological systems (e.g. case, verb agreement) and intonational contours. In comprehension form needs to be mapped to function, and in production function needs to be mapped onto form. While the number of surface forms is limited, the number of relational structures generated by lexical items is potentially infinite creating competition among the functional categories.

(12a) The ball hits the boys.

(12b) The ball hits the boy.

In sentence (12a) the lexical-semantic cue 'animacy' activates *the boys* as the agent. However, *the ball* receives activation as the agent by two other cues: the word order cue 'preverbal position' and the morphological cue 'verb agreement'. Word order and verb agreement form a coalition and win over animacy, so the final interpretation selects *the ball* as the agent. In (12b) the verb agreement cue is uninformative as it agrees with both NPs and there is only competition between animacy and word order. Since word order is the stronger cue in English, (12b) is also interpreted as an SVO sentence. Winning the competition for a cue strengthens the connection between the winning function and the cue, while the connection is weakened for the losing functions. This competition process is what gave the name to the model.

Cues in the CM have different properties that play an important role in its processing account and will be explained below. Cue **strength** is determined by success in the competition, and is highly language-specific as it reflects how often a particular cue wins the competition for a specific task, such as identifying an agent. If animacy often wins the competition for the agency assignment, its cue strength is high. A strong predictor of cue strength is cue **validity**. Cue validity is the combination of how frequent a particular cue is in the input (cue **availability**) and how consistently the cue maps to a particular form (cue

reliability). When a cue is always present during a task it has the highest possible level of availability. When a certain cue always allows the correct conclusion it has peak reliability. Conflict validity is a subcase of cue validity and represents the validity of a cue in conflict sentences, i.e. sentences in which cues lead to different results, and is therefore very important in sentence processing. In (12b) for example, the preverbal position has higher conflict validity than animacy. It wins the competition and determines how the sentenced is processed, i.e. as an SVO sentence. Cues can be low in availability, i.e. they are used very infrequently, but if they are used in competition with other cues, they have maximal reliability and therefore also high conflict validity.

Another concept central to sentence processing in the CM is the idea of cue **cost**. It reflects processing limitations caused by difficulties in cue perception and demands on working memory. Cue **perceivability** refers to the ease of detecting a cue in the input, and low perceivability is often reflected in a delayed acquisition of said cue. Cue **assignability** reflects the amount of information that needs to be processed for a particular cue. Cues that put less strain on working memory are high in assignability, while those that are more taxing have low assignability. Cue perceivability and assignability also influence cue reliability, as speakers will rely less on a cue that comes with a high cost.

Another cue property that reflects demands on working memory is the division into **local** and **global** cues. Local cues, i.e. animacy or case marking, require less information processing as they can be processed locally within a single word and do not need to be held in working memory. Global cues, i.e. word order, suprasegmental stress patterns or morphological agreement, require more information processing as they involve several words and more information that must be held in working memory. It has been suggested that languages could be divided according to their use of local and global cues.

By means of the agent identification task (which is covered more extensively in Section 2.4.3), researchers working within the CM framework aim to determine language-specific cue hierarchies that are at play during online sentence processing. Cross-linguistic differences emerge at a qualitative and quantitative level as languages make use of different types of cues and also rely differently on the same cue. For example, the word order cue is maximally reliable in English (McDonald, 1987), while it plays only a minor role in the

German cue hierarchy (MacWhinney et al., 1984). The CM makes the following predictions for online sentence processing behavior. It assumes a negative relationship between cue strength and online processing: the stronger a cue is, the faster it is processed. Converging cues further facilitate reaction times, while competing cues inhibit reaction times. In situations in which there is competition or convergence between a number of cues, effects of weaker cues on reaction times can be erased by the dominant influence of a single very strong cue. Depending on the settings of individual cues, a non-canonical word order does not necessarily have to show a great processing disadvantage compared to a canonical one. If the strongest cues within a language's hierarchy favor a non-canonical interpretation (e.g. unambiguous case marking in German), it should be processed more easily than in a situation in which there is competition between the cues and a canonical and non-canonical interpretation receive support.

The changes in cue strength that speakers experience during the acquisition process according to the CM are typical of **experience-based processing** models. The associated patterns and interactions between available sources of information in these types of models are often shaped by language use and experience with specific structures. Abstract principles as proposed by syntax-first models are less important to the parser. Frequencies of constructions and probabilities of co-occurrence play a central role to the parser as the parsing mechanism is flexible and adjusts itself based on these frequencies. Corpus studies are often used to assess the relative frequencies of individual constructions. Neural networks that use previous exposure to learn grammatical patterns are developed based on experience-based theories. For any parse, the parser is initially biased towards the most frequent structural interpretation and later adjusts this parse in accordance with the incoming information. The ambiguous Norwegian sentence in (10), would be parsed as an SVO sentence, because of the higher frequency of SVO orders compared to OVS orders. The parser would never a priori assume an OVS structure, especially in the absence of supporting case marking.

(10) Mannen kysser kvinnen
'The man kisses the woman./The man, the woman kisses.'

As experience-based parsing models draw heavily on frequency their predictions regarding non-canonical word orders are fairly clear. Non-canonical orders are less frequent than canonical ones and should therefore be dispreferred in parsing. For L2 speakers, experience-based models would also predict a change of processing preferences over time, as the speakers start with equal frequencies for all structures (zero) and gain more experience with canonical structures compared to non-canonical structures while gaining more experience with the L2 language overall.

A challenge for experience-based processing models comes from experimental findings that show divergences between corpus frequencies and parsing preferences (e.g. Desmet, De Baecke, Drieghe, Brysbaert, & Vonk, 2006) or between corpus frequencies and acceptability ratings (e.g. Kempen & Harbusch, 2005). Critics of experience-based processing models suggest a mixture of experience-based and memory-based accounts to explain these results (Boston, Hale, Vasishth, & Kliegl, 2011; Fedorenko, Woodbury, & Gibson, 2013; Lewis, Vasishth, & Van Dyke, 2006). Another problematic issue from an SLA perspective is the basis of the frequency counts. While written and spoken corpora of native speakers might give a decent approximation of the experiences of a native speaker, they are less suitable to reflect the experience of non-native speakers. Learner corpora that reflect text book frequencies and actual usage by learners at different stages of proficiency are needed in this case. The number of these corpora is rising, but given the variability in L2 input depending, for example, on teaching materials in non-immersive settings and individual experience in immersive settings, the accuracy of their data remains problematic.

2.1.3 Memory-based processing accounts

Alternative accounts that assign a less prominent role to syntax and focus instead on the ability of the parser to maintain intermediate parsing results are memory-based accounts. They can broadly be divided into **similarity-based accounts** (Lewis & Vasishth, 2005) and **distance-based accounts** (Gibson, 1998; Warren & Gibson, 2002). Both types of accounts assume that the representation of elements in memory decays over time. Non-local dependencies incur higher processing costs because the first element needs to be retrieved from working memory when the second element of the dependency is encountered. In example

(13) below, the main verb *looked* has to be retrieved from memory when the particle *up* is encountered as they form a non-local dependency.

(13) The secretary **looked**$_1$ the number of the chancellor **up**$_2$ this morning.

Decayed representations make this retrieval more challenging. The two approaches differ with regard to whether decay is more influenced by the pure distance or by the number of similar elements stored in working memory competing for selection. Similarity-based accounts predict higher processing difficulty the more similar the constituents in memory are. Their similarity makes it more difficult to differentiate them in memory and causes interference during retrieval. Similarity in this case does not necessarily mean orthographic or phonological overlap, but rather structural similarity. Distance-based accounts predict generally higher processing costs the greater the distance between the constituent and its integration site.

Both accounts are interesting for the experiments in this thesis as non-canonical orders often contain displaced elements and are therefore potentially more taxing for working memory. Similarity-based accounts might be better suited to explain the processing and comprehension of the Norwegian object-topicalizations (Study 1). In this study there are two NPs without morphological case marking that are potential agents when the OVS structure requires a revision and when the comprehension question requires the selection of an agent. The distance-based account is relevant to the manipulation in the experiment on Norwegian particle verbs (Study 3) as the length of the object is manipulated and increases the distance between the verb and its particle. The crucial point is that the dependency between verb and particle is correctly identified to initiate the search for the verb and its retrieval from memory which should incur higher processing costs compared to a local dependency. If the dependency is not recognized, no memory retrieval takes place and no memory effects can be found.

2.1.4 Models assuming incomplete parsing

All of the above accounts assume that processing in a native language results in representations that are complete, detailed and above all accurate reflections of the input. However, there are accounts in the literature, that suggest that even native processing can lack syntactic depth and be restricted to a **'good-enough'**

comprehension of the linguistic material (e.g. Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007). Under the time pressure of live discourse, the parser does just enough to reach a meaningful interpretation that is sufficient for the communicative task at hand. As reanalysis and the evaluation of all available sources of information take time that is not always available in discourse, the syntactic representation of a given sentence might not be detailed enough to represent important syntactic distinctions. Evidence for this approach comes from the persistence of initial misinterpretations after a reanalysis of garden path sentences like (14) below (Christianson, Hollingworth, Halliwell, & Ferreira, 2001) and from misunderstandings of unambiguous, non-canonical sentences such as the implausible passive in (15a) (Ferreira, 2003):

(14) While Anna dressed the baby played in the crib.

A detailed syntactic analysis of sentence (14) would require a reanalysis of *the baby* from being the object of *dressed* to being the subject of *played*. Christianson et al. found that speakers did not completely abandon this first analysis: they still gave a positive answer to the comprehension question asking whether Anna dressed the baby. A control group that read a non-garden path version with a comma after *dressed*, did not show this effect and gave the correct negative answer to the same comprehension question.

(15a) The dog was bitten by the man.

(15b) The man bites the dog.

(15c) The man was bitten by the dog.

(15d) The dog bites the man.

Ferreira (2003) presented readers with the implausible passive sentence in (15a), the implausible active sentence (15b), and a passive and an active version of the plausible scenario (15c,d). She found that unlike implausible active sentences, implausible passive sentences such as (15a) were rated as plausible more than 25% of the time. In an agent identification task, participants also showed an elevated error rate only for implausible passive sentences, but not for any of the other three types of sentences. Ferreira explained this effect through the use of semantics and world knowledge that overrule the correct syntactic representation. Ferreira and colleagues claim that a lack of reinforcement of the computed accurate linguistic representation can lead to a good-enough interpretation. This reinforcement can happen through context, world knowledge or schemas (defined as 'general frameworks used to organize details on the basis of previous experience', Ferreira, Bailey, & Ferraro, 2002:13) that are retrieved from long-term memory. A second source for good-enough interpretations is interfering information from previous incorrect interpretations or schemas, if it is at conflict with the result of the syntactic interpretation and cannot be inhibited enough.

Good enough processing does not necessarily have a negative effect on the processing of non-canonical word orders. This type of processing is mainly detected by abnormal accuracy scores for comprehension questions. Word order variations that do not cause a change of sentence content, as is the case for the scrambled sentences investigated in Study 2 and the particle placement alternation investigated in Study 3, are therefore open to good-enough processing as the order of the elements does not affect the response. The order manipulation of Study 1, however, would reveal good-enough processing when object-first sentences would be interpreted as subject-first sentences.

This section provided an overview of some of the current language processing models and tried to illustrate the different predictions these models make with regard to non-canonical word orders. I adopt a weak syntax-first model as my processing model in this thesis, without necessarily claiming its accuracy in describing the results of the experiments. I chose this model because the majority of the research on scrambling and subject/object ambiguities has been carried out based on syntax-first models and it can also be applied equally well to all experiments conducted for this thesis and their different experimental manipulations. Additional models will be considered when the experimental manipulation allows different processing predictions derived from these models compared to syntax-first models.

2.2 Processing models in L2 research

Learning a foreign language as an adult or even as a teenager is much harder than acquiring any language from birth or early childhood. The reasons for and consequences of this difference are at the heart of L2 research. Research mainly from the last 25 years has found that unlike L1 acquisition, L2 acquisition is characterized by great variability regarding its developmental sequence and final outcome, including the representation of specific grammatical rules. While L2 speakers might show the same knowledge of a rule, e.g. the availability of object topicalization in Norwegian, and perform similarly in an offline task, it is unclear whether the underlying syntactic representation of the object topicalization structure is the same for all L2 speakers. L2 acquisition is also more influenced by the learner's age, the quality and quantity of instruction and individual variables such as motivation and aptitude (see Slabakova, 2009 for an overview; Bley-Vroman, 1990). Despite its uncertain outcome, millions of people take on the challenge of learning a foreign language and their successes or failures can shed light onto our general understanding of language acquisition and processing.

When comparing L2 processing to L1 processing, especially in the domain of syntax, some areas seem to be harder to acquire for L2 speakers than other areas, resulting in prominent differences between L1 and L2 behavior. Elements of complex syntax, such as non-local dependencies, sometimes reveal non-native processing patterns even for highly proficient L2 speakers , while native-like processing tends to be found in some local dependencies, such as gender agreement (Foucart & Frenck-Mestre, 2012; Sagarra & Herschensohn, 2011). A variety of factors have emerged from the literature that serve as possible culprits for the non-nativeness of L2 processing, such as:

- a lack of relevant grammatical knowledge, i.e. the speaker has not (yet) acquired the construction in question and therefore has a non-target-like interlanguage grammar;
- an influence of the L1 through the use of L1-specific processing routines or a more general effect of the L1 lexicon and grammar;
- limitations of the cognitive resources available to the L2 speaker, causing less efficient integration of information or a delay in processing;
- maturational changes that do not allow the same type of acquisition as for the native language (critical period effects);
- less exposure to the L2 resulting in less automaticity in processing

Several theories and approaches can be distinguished that integrate and highlight these factors in different ways. While the statistical null hypothesis in second language research always assumes no difference between native and non-native speakers, the theories of L2 processing are aimed at explaining the source of the differences that have been found between native and non-native processing. A multitude of theories have emerged in the past 25 years and I selected those theories that are best suited to address the experimental manipulations used in this thesis. This section does not intend to give an exhaustive overview of all available L2 processing models.

2.2.1 Models assuming different representations in L1 and L2

One camp of researchers assumes that the differences between L1 and L2 speakers are **qualitative**, i.e. L2 speakers use different processing routines and have different linguistic representations than L1 speakers. The source of this qualitative difference varies with the hypotheses and ranges from the unavailability of access to Universal Grammar (UG, a set of rules in a language-specific learning capacity proposed by Chomsky, see e.g. Berwick, Pietroski, Yankama, & Chomsky, 2011) due to critical period effects, to the use of different neurological pathways. One of the earliest hypotheses in this field is the **Fundamental Difference Hypothesis** proposed by Bley-Vroman (1990) that was developed on the base of the abovementioned observed differences in acquisition and outcome between child L1 learners and adult L2 learners. It assumes a critical period for language acquisition (usually around puberty) that is responsible for the qualitative differences between L1 and L2 grammatical representations and processing. From these basic assumptions newer theories have been derived with more specific claims and predictions.

Working within the **Competition Model** (CM) framework discussed in the previous chapter, Kilborn & Ito (1989) suggested that L1 appropriate processing strategies and cue hierarchies can carry over into L2 processing. An English native speaker learning German would predominantly use the word order cue as it is the dominant cue in his native language, while an Italian learner would tend to use verb agreement. They also suggested a stronger, more central role for semantics in L2 processing compared to syntax. L2 learners may attend to cues that seem global and robust and built up their processing strategies in a linear fashion following cue validity. Also taking a CM perspective, McDonald (1989) proposed that L2 speakers map cues to categories based on a learning-on-error mechanism: cue weights are only adjusted after a miscategorization and not reinforced by accurate categorization. Easily detectable cues with high overall validity should therefore be acquired first. McDonald also suggested a gradual shift from L1 conflict validities to L2 conflict validities.

The **Shallow Structure Hypothesis** (SSH) proposed by Clahsen & Felser (2006a, 2006b, 2006c) builds up on the assumption of two grammatical representations in native speakers that is known from dual-route models (see Ullman, 2001a, 2001b, 2001c) and the good enough processing approach (see Section 2.1.4). One is a 'deep', complex, fully specified representation that is fed by the grammar, the other one is a 'shallow', rough representation that is calculated based on lexical-semantic and pragmatic information and statistical patterns. In native speakers, these routes serve different purposes: the shallow route calculates a first rough 'draft', while the deep route provides further confirmation for this draft. The SSH postulates that even highly-advanced L2 speakers are largely limited to the shallow route and rely more on nonstructural information such as lexical-semantic and pragmatic information to compute an incremental parse of the input, while L1 speakers' processing is more determined by the syntactic phrase structure of the input. Clahsen & Felser (2006a) assume that L2 speakers can access the same general parsing mechanisms they use in their L1, but that their application may be restricted due to an incomplete or insufficiently automatized L2 grammar. This difference in processing strategies can lead L2 speakers to perform in a non-native way in contexts that require more elaborate or 'deep' syntactic representations for successful parsing like cases of ambiguity resolution or non-local dependencies. The L2 speaker's internal representation of these structures may lack syntactic detail, i.e. a complex hierarchical structure and abstract elements such as movement traces. In the absence of such detail, the parsing commitments of L2 speakers might be weaker than those of an L1 speaker and no indication of reanalysis might be found. A stronger reliance on lexical-semantic and pragmatic information makes L2 speakers also more error-prone whenever the syntactic structure goes against the usual lexical-semantic preferences, e.g. inanimate subjects with animate objects instead of animate subjects paired with inanimate objects. While the syntactic part of the SSH in its original form from 2006 mainly considered non-local domains, Jin, Åfarli, & van Dommelen (2007) found shallow processing in highly-proficient L2 speakers of Norwegian also in the local domain of DP internal agreement. Speakers who had shown native-like behavior in the production of DPs were not able to use agreement cues in an auditory perception task. This behavior was independent of their L1 background as three different L2 groups were tested (English, Chinese and Romance) and no consistent evidence for transfer was found. The evidence from the morphological part of the SSH had already found shallow morphological processing in local domains.

Even though the CM framework and the SSH do not seem reconcilable at first, because they make very different assumptions regarding syntactic processing in general, their assumptions regarding L2 processing are very similar: Both emphasize the stronger use of semantics over syntax in L2 processing. Differences between semantics-driven and syntax-driven processing are more likely to surface in Studies 1 and 2, as Study 3 and 4 provide little semantic information that could lead to processing differences between the experimental conditions.

The **Interface Hypothesis** developed by Sorace and colleagues (Sorace & Serratrice, 2009; Sorace, 2011; Tsimpli & Sorace, 2006) focuses on the syntax-discourse interface dimension. The basic assumption of this hypothesis is that features that are relevant to the syntax-discourse interface are especially problematic in L2, e.g. the availability of pro-drop in Italian and Spanish that depends on the discourse (Domínguez & Arche, 2008; Sorace, 2007). It attempts to explain patterns of optionality and variation that have been found in highly advanced L2 speakers, such as the inappropriate use of pro-drop. Interface structures are assumed to be more costly in processing as they require knowledge of several types of information (e.g. syntactic and discourse-pragmatic knowledge) and their coordination. They are hard to acquire for L2 speakers as the variation encountered is not a matter of grammaticality, but of acceptability in certain discourse situations. The L2 interlanguage then shows signs of instability and is prone to continued L1 interference effects. Study 2 on German midfield scrambling and Study 3 on Norwegian particle placement investigated structures that show optional variations of word order that are supposedly caused by an influence of discourse and do not constitute a grammatical/ungrammatical dichotomy, but rather an acceptability gradience. Based on the Interface Hypothesis, L2 participants in these two studies would be less likely to show native-like behavior than the L2 group in Studies 1 and 4 that investigated phenomena that are less influenced by discourse and in the case of Study 4 show a clear grammaticality distinction.

2.2.2 Models assuming similar representations in L1 and L2

Another group of researchers argues that any observed differences between L1 and L2 processing are caused by a **quantitative difference** between the two groups. Grammatical representations and the processing architecture are not fundamentally different in native and non-native speakers and not affected by a critical period. They propose that cognitive factors such as limitations of working memory have a bigger impact on L2 processing than on L1 processing, such that L2 speakers cannot access the required knowledge fast enough to ensure smooth processing.

The **Fundamental Identity Hypothesis** (FIH) proposed by Hopp (2007) argues that there is no qualitative difference between the grammatical representations and the processing behavior in L1 and L2. It also rejects the idea of a critical period for L2 acquisition. Any observed differences are instead related to factors of L2 acquisition that are not influenced by a critical period, such as L1 transfer or performance factors, i.e. computational limitations. Rather than predicting statistically identical performance for L2 and L2 speakers, the FIH predicts the same overall processing pattern with possibly slower reaction times in the L2 group. Individual differences between L2 speakers in a variety of cognitive functions have also been investigated with regards to their influence on L2 processing performance (Hopp, 2015)

Experimental evidence has been used to support either of the two views and to establish the abovementioned theories, but the results are generally very mixed and do not allow us to rule out one approach conclusively.

2.2.3 A word on transfer

A highly contested topic in SLA research is the question of L1 transfer during acquisition and the role of the L1 during sentence processing in the L2. My own research cannot contribute much to this discussion as only one experimental L2 group was used (L1 German for Studies 1 and 3 on Norwegian, L1 Slavic for Studies 2 and 4 on German). Due to the difference in typological proximity between the L1s and the L2s, the reader may ask if some results could be explained by transfer, especially for the closely-related languages Norwegian and German. I will therefore shortly introduce the main points made by L2 processing theories regarding transfer.

Transfer can take different forms in the theoretical literature: full transfer, partial transfer, or no transfer. Full transfer models such as the Competition Model or the **Full Transfer/Full Access model** (Schwartz & Sprouse, 1996) assume that at the very first contact with a foreign language, everything is transferred from the L1. In the case of the CM this entails the transfer of cues and their individual values with regards to cue strength, conflict validity etc. In these theories the initial transfer occurs whether it is useful or not. Positive transfer, i.e. the transfer of identical structures from the L1 to the L2, and negative transfer, i.e. the transfer of structures that are unique to the L1, are both possible. Partial transfer models like the **Developmentally Moderated Transfer Hypothesis** (Pienemann & Håkansson, 2007) assume that the developmental state of the interlanguage system determines the timing of this transfer. If a structure cannot yet be processed by the interlanguage system, there is no transfer. According to this hypothesis there is only positive transfer as transfer from the L1 only occurs when it is useful and results in grammatical structures in the L2.

When experimental results of an L2 group look similar to the expected pattern in their native language, L1 transfer is often quickly claimed to be the source. The problem with this claim is that many theories only assume transfer during the very initial stages of contact with the L2 and a gradual development towards a native-like L2 end state. They would not predict 1:1 transfer effects in highly advanced speakers like the L2 speakers investigated in this thesis. Even if these speakers have not yet reached the end state of their L2 development, their interlanguage grammar should differ from their L1 grammar because of their experience with the L2. The argument that L2 speakers might not have encountered a certain structure and therefore show transfer effects seems untenable given their experience and proficiency.

Another problematic point that weakens the L1 transfer argument is the experimental setup: In most L2 studies only one L2 group is tested, although there are notable exceptions (e.g. Gerth, Otto, Felser, & Nam, 2015; Hopp, 2005). If this single L2 group then shows a behavior that looks like it could be caused by L1 transfer, this cannot be verified as it might also be the 'standard' L2 behavior that is unrelated to the L1 background and could be found in any L2 group. The studies in this thesis also used only one L2 group and I will therefore not assign a lot of space to the discussion of transfer effects as any assumptions are highly speculative in the absence of a second L2 group.

2.3 Theoretical approaches to word order phenomena

The studies in this thesis investigated word order variations that contrasted canonical and non-canonical orders, such as the non-canonical placement of objects in a topicalized positions and the alternation of objects in ditransitive sentences or of the direct object, and the placement of the particle in particle verbs. Theoretical approaches concerned with the generation of canonical and non-canonical word orders will be discussed in this chapter and the main terminology that is relevant for the following experiments will be introduced.

2.3.1 Derivational approaches vs. base generation

One main theoretical divide in the research on word order is between derivational approaches and base-generation approaches. **Derivational approaches** such as Chomsky's **Principle and Parameters Theory** (PPT) (Chomsky, 1981) assume one 'basic' word order from which other non-basic orders can be derived via movement. Moved constituents leave behind a trace or copy (Chomsky, 1995) at their original syntactic position. In the wh-question in (16), the wh-element has moved to the sentence-initial position and left a trace at its original position in the VP.

(16) Who$_i$ did you call t$_i$ on your birthday?

This trace is a place-holder that is syntactically active, but phonetically empty and has the same grammatical features as the moved constituent. Theory-neutral terms used in psycholinguistics that describe a constituent and the position where it can be integrated are **filler** and **gap** (Fodor, 1989). The relation between those two elements is then called a **filler-gap dependency**.

There are several types of movement and I will briefly explain two distinctions that will be relevant for the experiments reported in this thesis. One is the distinction between **A-movement** and **A'(A-bar) movement**. Argument-movement (A-movement) involves the displacement of a phrase to a position that assigns a fixed grammatical function and commonly occurs in passivization and with unaccusatives.

(17a) I read Chomsky's book on Minimalism.

(17b) Chomsky's book on Minimalism$_i$ was read t_i by me.

In the passive sentence (17b) the DP *Chomsky's book on Minimalism* receives its thematic role of patient in its original position in VP. It then undergoes A-movement to the sentence-initial position in IP where it receives nominative case. A-bar or non-argument movement on the other hand moves a phrase into a position that does not assign a fixed grammatical function. Wh-movement in English as in (16) is an example of A-bar movement. The wh-element is moved from a subject or object position to a pre-verbal position and despite the movement, the wh-element retains its original case.

The second distinction is the one between **phrasal movement** and **head movement**. This distinction is quite straightforward. Phrasal movement concerns the movement of a head with all its dependents, thereby moving the entire phrase. Sentences (16) and (17b) are instances of phrasal movement as in both cases the whole DP has moved. Head movement is restricted to the movement of a phrasal head alone, leaving behind its dependents. This occurs for example in subject-auxiliary inversion in yes/no-questions as (18b) below.

(18a) You have seen the latest Bond movie.

(18b) Have$_i$ you t_i seen the latest Bond movie?

The distinction between head and phrasal movement is often used in accounts of particle placement in particle verbs in order to account for stranded particles and those particles that move along with the verb (see Wurmbrand, 2000; Zeller, 2002).

Entrenched in Chomsky's theory of transformational grammar is also his case theory. Case expresses the function of an NP in the sentence and can be influenced by NP movement. Case is always assigned on a syntactic level and some languages, such as German or the Slavic languages, also express case through morphological markers. The exact way how **morphological case** is expressed varies with the language, and morphological markers can for example be found on determiners, adjectives, pronouns or nouns. Table 2.1 below compares the expression *the young man* as subject, direct object and indirect object in English, German and Russian. In this example, English shows no overt case marking. German only shows overt case marking on the determiner and the adjective and the case marking of the adjective is identical in accusative and dative case in this paradigm. Russian has different case marking on the adjective and the noun for both object cases.

	English	German	Russian
Subject (nominative)	the young man	der junge Mann	molodoj čelovek
DO (accusative)	the young man	de**n** junge**n** Mann	molod**ogo** čelovek**a**
IO (dative)	to the young man	de**m** junge**n** Mann	molod**omu** čelovek**u**

Table 2.1 Comparison of morphological case marking in English, German and Russian

Abstract or **structural case** is assigned in the syntactic representation. The exact structural position in which for example nominative case is assigned can vary between languages. Nominative case in English is assigned in SpecIP. NPs that originate in SpecVP need to move to SpecIP to receive nominative case. NPs that are complements of VP are automatically assigned accusative case (see schematic in figure 2.1). Figure 2.1 only illustrates structural case assignment in English, it does not illustrate the verb movement necessary to account for tense.

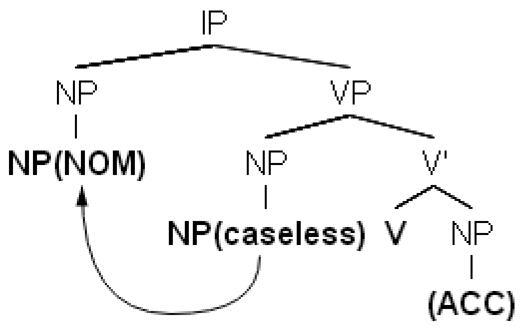

Figure 2.1 Schematic of structural case assignment in English

As structural case is assigned based on specific positions in the syntactic tree, it is not preserved under A-movement. As can be seen in examples (17a + b), the object of sentence (17a) moves from its original position in VP to SpecIP under passivization (17b) and therefore receives subject case.

Lexical case is a **non-structural case**, meaning it is not automatically assigned based on the syntactic structure, but is assigned as a property of specific verbs (see Woolford, 2006 for an additional distinction between lexical and inherent case). As lexical case is licensed by certain lexical heads, it is highly idiosyncratic and is preserved under A-movement. Dative case is a lexical case that is assigned by ditransitive and unaccusative verbs. The German example in (19a +b) illustrates the preservation of dative case under passivization.

(19a) Die hübsche Frau gibt dem jungen Mann einen Kuss.
the$_{NOM}$ pretty$_{NOM}$ woman gives the$_{DAT}$ young$_{DAT}$ man a$_{ACC}$ kiss
'The pretty woman gives the young man a kiss.'

(19b) Dem jungen Mann wird ein Kuss gegeben von der hübschen Frau.
the$_{DAT}$ young$_{DAT}$ man PASS a$_{ACC}$ kiss given of the$_{DAT}$ pretty$_{DAT}$ woman
'The young man is given a kiss by the pretty woman.'

The movement of constituents and the presence of traces are not without consequences for the parser as the association of a filler and its gap needs to be resolved during online processing. First, the parser has to identify the displaced element as a filler, and depending on the shape of the element this identification can already be difficult. Due to case ambiguity the German NP *die Frau* 'the woman' can either be a subject or an accusative object. When the parser encounters this NP in a sentence-initial position, case marking is not informative and the parser likely interprets the NP as a subject and not as a moved object. Other sentence-initial NPs are less problematic to interpret on the first pass. For instance, an unambiguously case-marked NP like *den Mann* 'the man (accusative)' immediately signals a filler-gap dependency when encountered in a sentence-initial position. Once the parser has identified a filler, it has to be kept in short-term memory until the parser finds a suitable gap, putting a strain on processing resources. As the gap has no surface manifestation its identification can be difficult and needs to be inferred from the syntactic structure (see Hawkins, 1999). Inferred gaps can also be filled by other material and the parser then has to continue its search for another gap.

Approaches that challenge Chomsky's ideas of derivation and movement are **non-derivational** or **base-generation approaches** that do not entail the movement of constituents. Instead all orders are the result of the free generation of constituents and are therefore equally 'basic'. As there is no movement in these approaches, there are also no traces. Trace-free theories are common in lexicalist frameworks such as **Head-driven Phrase-Structure Grammar** (HPSG; Pollard & Sag, 1994) or **Lexical Functional Grammar** (LFG; Bresnan, 2001). The absence of traces does not mean that the syntactic dependency has also been abandoned. HPSG for example assumes feature sharing between the locally absent dependent and its subcategorizing head.

I adopt a movement account in this thesis as Study 2 was designed to test different predictions of movement and base-generation accounts. However, this is mainly for descriptive purposes only: the presence of a reading time difference in the self-paced reading tasks alone is no evidence for a movement account as these differences could also be caused by frequency effects for two base-generated structures. It is therefore difficult to tease apart derivational and non-derivational accounts with the methods and materials used in this thesis.

2.3.2 Topicalization vs. scrambling

Part I of this thesis, presented in Chapters 3 to 5, is concerned with two types of discontinuity: topicalization and scrambling. As I assume a derivational framework for the purpose of this thesis, I assume that these continuities are caused by the movement of the object either to a sentence-initial position or within the center of the sentence (the midfield). Topicalized and scrambled structures have the same truth-conditional meaning as their non-topicalized or non-scrambled counterparts. The difference between the moved and the canonical sentences therefore lies in their focus structure. The canonical, unmarked order has no restrictions with regard to its potential focus structure, while non-canonical, marked orders have very specific focus requirements.

As the name already suggests, **topicalization** is used to establish a constituent as the topic of a sentence or clause. In the languages investigated in this thesis (Norwegian and German), this is achieved through the movement of the constituent to the sentence-initial position, as in examples (20a-c) in which the topicalized element has been printed in bold.

Canonical word order
(20a) Der Mann gibt der Frau das Geschenk.
the$_{NOM}$ man gives the$_{DAT}$ woman the.ACC present
'The man gives the woman the present.'

Topicalized indirect object
(20b) **Der Frau** gibt der Mann das Geschenk.

Topicalized direct object
(20c) **Das Geschenk** gibt der Mann der Frau.

Topicalization is also often used to test for constituency: expressions that can be topicalized together form a constituent. However, the exact syntactic representation of topicalized structures depends again on the syntactic theory assumed.

While topicalization is a wide-spread phenomenon across languages independent of the freedom of their word order, **scrambling** has been found to occur frequently in languages with a relatively free word order, e.g. German or Slavic languages. It is absent in languages with comparatively strict word order such as English. The term scrambling dates back to Ross (1986) and was first used in Ross' unpublished PhD thesis in 1967. It describes the clause-internal, non-obligatory movement of constituents to accommodate pragmatic needs and entails all possible non-canonical word orders in free word order languages. Scrambling applies to arguments of all categories and can be iterative (Haider, 2006).

The most widely investigated type of scrambling in German is midfield scrambling. The midfield is the section of the sentence between the finite and the non-finite verb in main clauses, or the section of the sentence between the subordinator and the finite verb in embedded clauses (see Chapter 5.1. for a more thorough introduction). Depending on the clause, it either affects only the objects (main clause, see 20d) or it can affect the subject and the objects (embedded clause, see 21a-f).

Scrambled version of (20a)
(20d) Der Mann gibt das Geschenk der Frau.

(21a) **Canonical embedded sentence**
...dass der Mann der Frau das Geschenk gibt.
...that the$_{\text{NOM}}$ man the$_{\text{DAT}}$ woman the$_{\text{ACC}}$ present gives
'...that the man gives the woman the present.'

DO scrambled across IO
(21b) ...dass der Mann das Geschenk der Frau gibt.

IO scrambled across S and DO
(21c) ...dass der Frau der Mann das Geschenk gibt.

IO and DO scrambled across S
(21d) ...dass der Frau das Geschenk der Mann gibt.

DO scrambled across S and IO
(21e) ...dass das Geschenk der Mann der Frau gibt.

DO and IO scrambled across S
(21f) ...dass das Geschenk der Frau der Mann gibt.

Whether a specific argument order is the result of scrambling depends on the verb involved as the neutral argument order is assumed to be encoded in the lexical entry of the verb. The canonical argument order for most verbs in German is agent > patient, however, for ergative and some psych-verbs it is patient > agent. An object > subject order alone is therefore no evidence of scrambling as it can also be the canonical order for certain verbs. Additionally, there is no 1:1 correspondence between agency and subjecthood, or between the patient role and objecthood. In general, experimental evidence points towards a preference of the subject > object in a variety of investigated languages and sentence contexts (see e.g. Bader, 2000 for German; Erdocia, Laka, Mestres-Missé, & Rodriguez-Fornells, 2009 for Basque), but the investigated languages usually have predominant SVO or SOV orders. Languages with predominant orders like OSV or OVS would probably show a preference for object > subject orders. Results from Basque, an ergative language, suggest that the preference for subject > object is not related to the case marking principles of a language. In nominative/accusative languages, it is usually the accusative and with it the object that receives the more complex case marking. In the Russian example in Table 2.1 the nominative has zero case marking and the accusative is marked by an –a ending on the noun. Here, the subject > object preference coincides with the less complex case marking preceding the more complex case marking. In ergative/absolutive languages like Basque the subject of a transitive clause has a more complex case marking, i.e. ergative case, than the object of that clause that is marked by absolutive case which is often expressed by zero marking. In this case the subject > object preference results in the more complex case marking preceding the less complex case marking. This general preference for subject-first sentences has consequences for online processing as the initially computed order will most likely be subject > object. Scrambling usually causes changes to the focus structure of the sentences as moved arguments tend to be focused. An exception to this rule are pronouns that tend to occur right after the complementizer in embedded sentences (the Wackernagel position), even if this causes an object-before-subject order. The movement of the pronouns in (22a-c) has no consequences for the focus structure as they occur in their canonical position.

(22a) …dass ihr der Mann das Geschenk gibt.
that her the$_{NOM}$ man the$_{ACC}$ present gives
(pronoun version of 21c)

(22b) …dass es der Mann der Frau gibt.
that it the$_{NOM}$ man the$_{DAT}$ woman gives
(pronoun version of 21e)

(22c) …dass es ihr der Mann gibt.
…that is her the$_{NOM}$ man gives
(pronoun version of 21f)

The analysis of scrambling within the derivational framework is not without controversy as for example scrambling in German shows properties of both A-bar-movement and A-movement (Webelhuth, 1989). Scrambled word orders have also been explained by base-generation approaches that often use linear-precedence rules to account for the different surface word orders. Fanselow (2001) advocated a base-generation of scrambling in A-positions, Bošković & Takahashi (1998) assumed base-generation in A-bar-positions. Kosta (2006) also takes a base-generation perspective on German and Czech scrambling and proposes IO > DO as the canonical object order in ditransitive sentences.

In their introduction to a volume on scrambling Corver & van Riemsdijk (1994:13) provide the following helpful schematic of approaches towards scrambling:

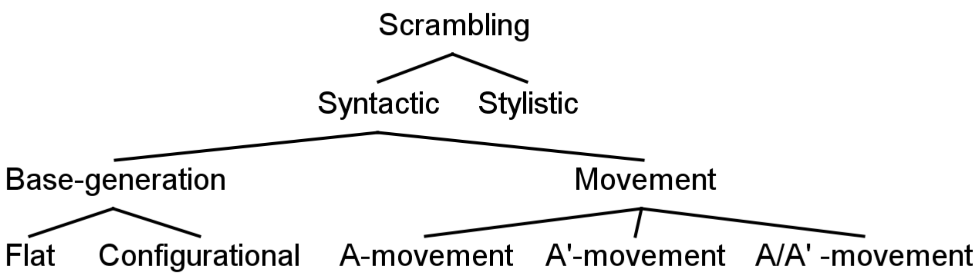

Figure 2.2 Schematic representation of approaches towards scrambling (Corver & van Riemsdijk, 1994)

2.4 Experimental methods

In this chapter I present an overview of the methods used in the experiments, how these three methods work, and justify why exactly these methods were used. Online and offline methods were chosen to complement each other. The online method of self-paced reading taps into real-time language processing and the unconscious application of syntactic rules and parsing strategies. Offline methods, like acceptability judgments or agent identification tasks, focus rather on the more conscious application of acquired syntactic rules and the final outcome of the parsing process. This section is especially aimed at readers that are less familiar with psycholinguistic experiments and want to know more about the experimental background than is reported in the materials and procedure sections of Chapters 4, 5, 6 and 7.

2.4.1 Self-paced reading (SPR)

Self-paced reading is one of the most popular experimental techniques in psycholinguistics and was used as the online method in all four studies of this thesis. Its first general application dates back to the 1970s (Aaronson & Scarborough, 1976; Mitchell & Green, 1978) and it has been used in second language research since the mid-1990s (Juffs & Harrington, 1996; see Papadopoulou, 2005 or Roberts, 2012 for reviews). The method is supposed to mimic normal reading (though critics of this technique question this assumption and point to the adoption of artificial reading strategies by participants; see Witzel, Witzel, & Forster, 2012 for a comparison of different online reading paradigms) and is mainly applied to investigate processing at the sentence-level and beyond. In SLA research, it has also been used as an online performance measure complementing grammaticality judgments and is considered to be a more implicit measure of grammatical knowledge as the time pressure disfavors the conscious application of explicit grammar rules. It also allows the experimenter to compare the processing behavior of native and non-native speakers (Jegerski, 2014; Juffs, 2001).

Unlike in eye tracking-during-reading paradigms which present the whole sentence or paragraph at once, SPR paradigms feed the experimental sentence to the reader in a piecemeal fashion. The reading time for each segment is then recorded separately and the length of the presentation is determined by the participant, therefore the term ‘self-paced’. There are several ways in which an experimental sentence can be presented. In the cumulative method, segments remain visible after the first presentation, and new segments are added until the entire sentence is presented on screen. In the noncumulative method, segments are masked again after the first presentation, when the participant continues to the next segment. The presentation of stimuli can either take place at the center of the screen with each new segment replacing the previous one or it can take place in a linear order in which the segments appear in a linear succession in the reading direction of the target language without spatial overlap. The noncumulative method with linear presentation of stimuli is also called the ‘moving window’ technique and is assumed to have the closest resemblance to natural reading (Just, Carpenter, & Woolley, 1982).

There are two main ways to present the segments: word-by-word or phrase-by-phrase. Phrase-by-phrase presentation is closer to natural reading, but also imposes the particular phrase grouping chosen by the experimenter upon the subject. Word-by-word presentation is less natural, but offers measurement on a more detailed level and can be converted into phrase-by-phrase reading times by summing up the reading times of individual words. For my experiments, I chose the moving window technique with word-by-word presentation. Figure 2.3 shows a screenshot from Experiment 2b showing the German noun *Richter* (‘judge’). At this point, the article has already been masked again by the three dashes preceding the noun.

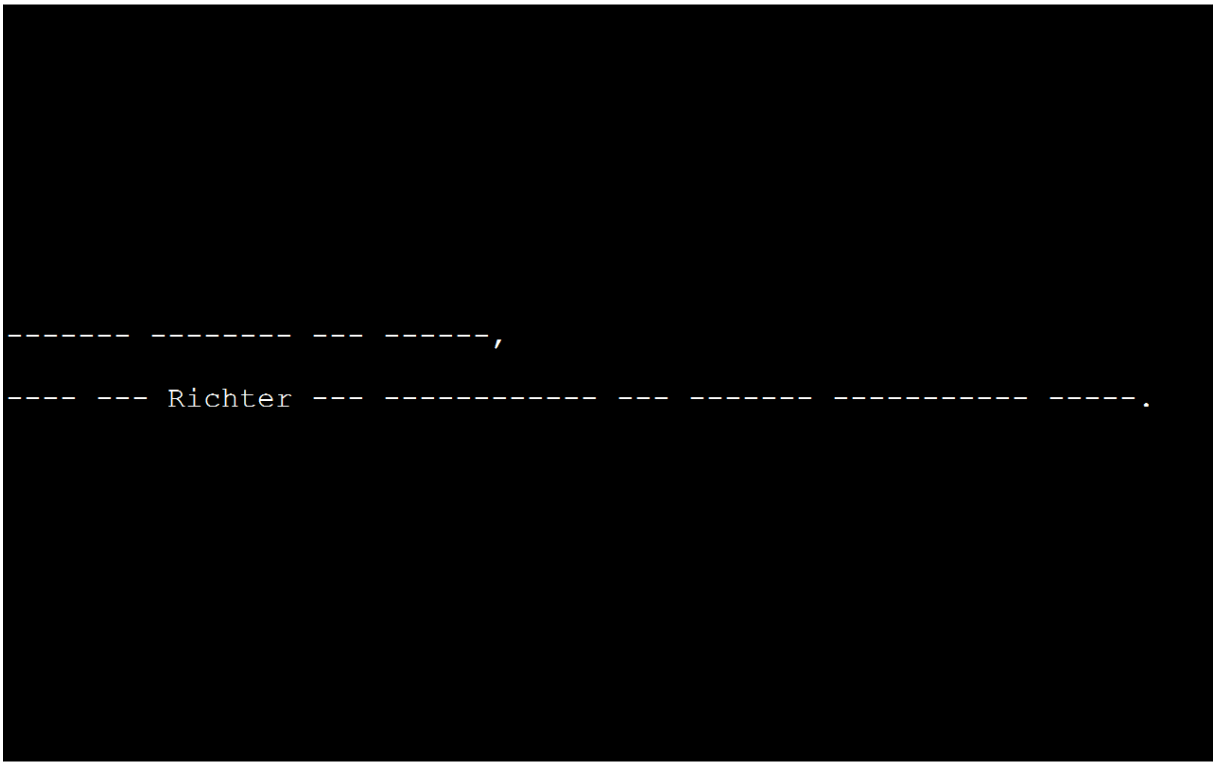

Figure 2.3 Screenshot from the SPR task in Experiment 2b

The basic rationale of SPR is that inferences about the online processing of language can be drawn from reading times (Just & Carpenter, 1980). Longer reading times relative to a baseline condition are assumed to signal difficulties in processing caused by an ungrammaticality, an ambiguity, or a revision of a previous interpretation, while faster reading times are assumed to signal facilitation. This line of reasoning is less clear for L2 speakers, as elevated reading times can either signal the same processing difficulties that L1 speakers experience or difficulties that are associated with L2-specific processing challenges (e.g. limited resources).

SPR has been used to investigate a wide range of phenomena that fall into three categories: ambiguities, anomalies, or discontinuous dependencies. All three of these phenomena are addressed in this thesis.

Structural ambiguities result from several possible syntactic analyses permitted by the grammar, and differences in reading times can be observed if the parser prefers one interpretation over another. Experimental items for this type of structure are always grammatical and the ambiguities can either be local ones that are resolved within the sentence or global ambiguities that remain unresolved (e.g. relative clause or PP clause attachment). In experiments both local and global ambiguities are resolved either towards the parser's preferred interpretation or towards its less preferred interpretation. The local ambiguity in (23a,b) exemplifies a preferred (23a) and a dispreferred (23b) resolution.

In local ambiguities this resolution is achieved through material provided in the sentence, whereas global ambiguities either remain unresolved or the reader must use contextual information to resolve the ambiguity (Pan & Felser, 2011). The preferred interpretation serves as the baseline for the reading times, while the less preferred interpretation should result in elevated reading times usually caused by the parser abandoning its preferred interpretation and subsequently reanalyzing the sentence. In sentences (23a,b) the two NPs have ambiguous case marking and can both be interpreted as either nominative or accusative, the disambiguation point is the verb-final auxiliary that either agrees with *die Frau* ('the woman') in singular or with *die Kinder* ('the children') in plural. (23b) is disambiguated towards the less preferred OVS interpretation and should show signs of reanalysis.

Preferred SVO interpretation
(23a) …dass die Frau die Kinder getroffen hat.
that the$_{\text{NOM/ACC}}$ woman the$_{\text{NOM/ACC}}$ children met has
…'that the woman has met the children.'

Dispreferred OVS interpretation
(23b) …dass die Frau die Kinder getroffen haben.
that the$_{\text{NOM/ACC}}$ woman the$_{\text{NOM/ACC}}$ children met have
…'that the children have met the woman.'

Anomalies can either be violations of grammar, or non-canonical permutations of word order, semantics or discourse (Jegerski, 2014). Two examples of SPR studies that investigated anomalies are Slioussar (2011) on scrambling in L1 Russian and VanPatten, Keating, & Leeser (2012) on several types of grammatical violations in L2 Spanish.

Another element of SPR experiments are additional tasks that follow after the stimuli. The main idea of these tasks is to ensure that participants pay attention to the task and actually process the sentences. They are also supposed to hinder participants from reflecting on the actual SPR task. Acceptability judgments and comprehension questions are the most frequently used types of tasks. In my experiments, I used comprehension questions throughout since the participants also completed offline acceptability judgments and only one SPR experiment actually contained grammatical violations.

My experiments targeted four different groups, L1 speakers of Norwegian, L2 speakers of Norwegian, L1 speakers of German and L2 speakers of German. These groups were spread over three different geographical locations (one in Norway and two in Germany). Since SPR is an experimental method that allows a lot of flexibility in testing locations and requires no highly specialized equipment, I chose this technique over for example eye tracking that could have been an alternative technique for investigating the same phenomena. I also wanted to control the presentation of the stimuli and make sure that participants could not skip shorter elements, like articles or particles, which are an essential part of the manipulation in three out of four of my studies. Self-paced reading was the best method to reconcile the experimental and the geographical needs of this thesis.

2.4.2 Acceptability ratings

Acceptability ratings have been used to assess speakers' introspective judgments of sentences since the 1950s, gaining major importance with the advent of generative linguistics introduced by Chomsky (1957). In theoretical linguistics these judgments are used to evaluate and describe various phenomena in syntax and semantics. The weakness of this method in theoretical linguistics that continues to be criticized until today (Gibson & Fedorenko, 2013) is that the judgments in published papers are often provided by the authors themselves, linguistic colleagues, or the first available informant. This non-experimental method of evaluating sentences involves only a very small set of participants and items. The selection of particular words can cause idiosyncratic effects on judgments that would be different for the same syntactic phenomenon but with different words in the sentence. Cognitive biases can lead to judgments that are influenced by insight into the theoretical problem and the hypothesis or the aim of the single informant to please the experimenter. Even changes to the instructions or the context in which the sentence is presented are known to influence the acceptability ratings. Zervakis & Mazuka (2013) also showed that acceptability ratings are sensitive to repeated evaluation of the manipulation in question: ratings get more favorable over time, and experimental lists therefore need to be carefully constructed with items counterbalanced.

In psycholinguistics acceptability ratings are mainly used in a quantitative, experimental way and rating tasks are carefully constructed with balanced and randomized presentation lists, several items per condition, plenty of filler items and a sizeable group of participants. Given a large enough group and a reasonable number of sentences plus distractors, the results are normally systematic across speakers and correlate with other dependent measures. Acceptability judgment tasks designed this way can be used to measure both native and non-native speakers' assessment of a wide variety of phenomena. They can be used to evaluate the application of grammatical rules postulated by descriptive grammar as well as speakers' preferences for grammatical variations. As a standardized method, acceptability judgments have the additional advantage that they are a comparatively simple task that can be administered either as a pen-and-paper questionnaire or web-based. One example for a large-scale study involving an acceptability judgment task is Gibson, Piantadosi, & Fedorenko (2011).

There are several ways of setting up the rating scale for an acceptability rating with the simplest 'scale' being a binary judgment with only two points. They all have in common that the two extremes of the scale correspond to labels such as "acceptable/not acceptable" or "natural/unnatural" A quantitative version of the scale has either the form of a Likert scale, usually with 5 to 7 points, or it is a geometric scale in which acceptability is compared to a given referent sentence. The latter method is called magnitude estimation (Bard, Robertson, & Sorace, 1996). Ratings are usually given on sentences in isolation to avoid effects of semantics or pragmatics.

A downside to acceptability judgments is that the researcher usually does not get an insight into the reason of the decision unless a thinking-out-loud paradigm is employed in which the participant has to state his or her thoughts during the decision process. The use of irrelevant metalinguistic knowledge that is unrelated to the grammatical manipulation also remains obscure without a form of thinking-out-loud paradigm or the need to justify the decision.

In the acceptability ratings for this thesis I used a five-point Likert scale and the ends of the scale were marked with translation equivalents of "acceptable" and "not acceptable" in the language assessed in the task (i.e. either Norwegian or German). I also encouraged my participants to correct those sentences that they judged as unacceptable or at least mark the source of the unacceptability. Each item had an additional row for these corrections (see Figure 2.4).

Du skal vurdere setningene om de er akseptable eller ikke. Akseptabel betyr her at de leses naturlig i dagligdags språk. Du kan gjerne bruke hele skala. Eksempel: Anne leser ikke boken. -> akseptabel = 1 Boken være spennende. -> ikke akseptabel =5	**Skala**				
	akseptabel				**ikke akseptabel**
Terje lar seg skille fra Elisabeth.	1	2	3	4	5

Figure 2.4 First item of the Norwegian acceptability judgment task used in Experiment 3a

Superficially, grammaticality and acceptability judgments seem to be the same task: participants are asked to rate the goodness of a sentence. The difference lies in the possible distinctions that can be made. Grammaticality, especially in the way it is used in theoretical linguistics, usually allows only a binary option of 'grammatical' vs. 'ungrammatical'. Authors in theoretical linguistics mark ungrammatical sentences with *, and may communicate nuances of uncertainty about their judgment via any number of question marks. Acceptability is assumed to vary continuously as grammar is only one factor in the judgment along with other components. Fanselow & Frisch (2006) for example argue that acceptability judgments are influenced by the degree and type of processing difficulty that raters experience. Grammatical sentences that cause processing difficulties by forcing the rater to adapt a less preferred interpretation are rated as less acceptable than sentences without processing difficulties (e.g. object-initial vs. subject-initial sentences), while local ambiguities that allow a reasonable intermediate parse can make ungrammatical sentences feel less unacceptable as the local acceptability carries over into the overall acceptability. Being a continuous factor, acceptability also does not have absolute endpoints, i.e. even the most acceptable sentence will not receive a perfect rating, and there are no clear criteria for a cut-off point that turns an unacceptable sentence into an ungrammatical one.

I used acceptability ratings in this thesis to have an offline performance measure that could then be compared to the online reading results. This was most relevant for the L2 group as differences between online and offline behavior are common in L2 speakers. In the Norwegian experiment on particle verbs I also wanted to verify the rules stated by prescriptive grammar as they are formulated very vaguely.

2.4.3 Agent identification

The agent identification task is deeply rooted in the theoretical framework of the Competition Model (CM) previously discussed in Section 2.1.2. The task has been used since the CM framework was first established in the mid-80s, and is the main tool used within this framework. Agent identification was first used as a an auditory task in child L1 acquisition and was later also used in a written form in research on bilingualism and SLA. Over the course of the last 30 years, more than 20 languages have been investigated with this method.

In the version of the task used by MacWhinney and colleagues, participants are given simple transitive sentences consisting only of two nouns and a verb and are asked to determine the agent of the sentence, for example in a sentence like *The cheese eats the dog*. The agent identification task itself can be administered both in an offline and an online format. The offline format measures the final response of the participant, that is, whether the first or the second NP was identified as the agent. It can either be administered as a pen-and-paper questionnaire or in the case of non-literate participants in the form of auditorily-presented sentences followed by a picture selection task. The online format measures reaction times and participants are instructed to choose the agent as quickly as possible.

The method has two weaknesses. As all investigated cues (e.g. animacy, word order, case or verb agreement) need to be systematically varied to allow generalization as to their ranking, ungrammatical sentences are also presented to the participants. Most of the time only simple sentences are used and extrapolation to more complicated structures is difficult.

In my thesis, I used the pen-and-paper version of the agent identification task (reported in Section 4.3) and participants had to consider three cues: NP animacy, preverbal NP position, and the order of the main verb and the second NP. Figure 2.5 exemplifies the use of the agent identification task as used in Experiment 1a. I used more complex sentences than is the norm in this task, but all sentences used were grammatical. Sentence 1 in Figure 2.5 shows a globally ambiguous filler item and sentence 2 a locally ambiguous experimental item.

1.____ Eleven mater sauen. (The pupil feeds the sheep.)

a) **sauen** b) **eleven**

2.____ Prinsessen ville prinsen male. (The princess, the prince wants to paint.)

a) **prinsen** b) **prinsessen**

Figure 2.5 First two items from the agent identification task from Study 1. (English translation not present in task)

One reason for using this task in my thesis is that it has been used on a wide variety of languages before, meaning I would be able to compare my results to those of many previous studies and possibly establish a cue hierarchy for Norwegian native speakers. A comparison of the cue hierarchy found in the agent identification task and the online reading behavior found in the SPR task can check the assumptions of the CM regarding cue convergence and cue competition. This is especially relevant for the L2 group as this group might show diverging cue hierarchies in the two tasks.

2.5 Summary

The current chapter has equipped the reader with the general background knowledge needed to be able to benefit from the entire scope of this thesis. Section 2.1 introduced the wide range of processing models that are currently on the linguistic market. For the purpose of this thesis I assume a serial syntax-driven model of processing, but other models such as the Competition Model or memory-based models will be discussed when they seem fit to explain the data and are motivated by the experimental manipulation. Section 2.2 introduced the theories dedicated to L2 processing that mainly center around the question to what extent native (L1) and non-native (L2) processing and syntactic representations are identical, and whether L2 speakers can ultimately achieve native-like status. The main purpose of this thesis is to compare native and non-native processing, the null hypothesis being that there is no qualitative or quantitative difference between the two groups. Section 2.3 provided the reader with the theoretical background on word order variations in general and object placement in particular. Throughout this thesis, I adopt a movement-based analysis of the phenomena under investigation. Finally, Section 2.4 gave an overview of the experimental methods that were used in this thesis. All three methods used in this study have a long tradition in linguistic research in general and also in L2 research. If administered carefully, all three methods produce reliable results, even though no method is without controversies.

PART I

-

PROCESSING OF OBJECTS IN NON-CANONICAL POSITIONS

3 Introduction

This chapter is an introduction to Part I of this thesis that is concerned with the processing of topicalizations and scrambled structures. Section 3.1 provides the reader with a general background on non-canonical object orders relevant to the studies on Norwegian object topicalization (Chapter 4) and German midfield scrambling (Chapter 5). I present an overview of the ordering possibilities in both languages and how they can be described syntactically, leaving more specific background information on the individual phenomena for the respective experimental chapters. Section 3.2 contains a review of relevant past research on non-canonical object orders, focusing on research on languages closely related to the ones investigated. Studies on L1 and L2 processing are reviewed side-by-side. Section 3.2 is broadly divided into two subsections: one pertaining to subject/object ambiguities and other garden path phenomena, the other to scrambling and dative alternation. Although the division between the two subsections is somewhat arbitrary, this organization will help the reader identify the relevant research questions that have been addressed already and the gaps in our knowledge that still need to be addressed.

3.1 Non-canonical object orders – theory and processing

Languages differ with respect to the degree to which they allow argument orders to vary within a sentence. English, which has minimal morphological case marking and verb agreement, has a very rigid word order, while languages that have a richer morphosyntactic agreement structure, such as German or Russian, usually allow several linear orders of the subject, and the direct and indirect objects. It seems that irrespective of how flexible the word order of a language is, there is always one basic word order that is used in context-neutral situations. This order is also processed and read the fastest (see Erdocia, Laka, Mestres-Missé, & Rodriguez-Fornells, 2009 for Basque; Slioussar, 2011 for Russian). This basic word order is assumed to be directly generated from the grammar in movement and base-generation approaches alike. All other non-basic or non-canonical word orders that occur in marked contexts can either be derived from this basic word order though movement (Chomsky, 1981) or they are also assumed to be base-generated (Fanselow, 2001; see also Section 2.3.1)

Several factors have been put forward in the theoretical literature to account for changes to the order of arguments. Among them are thematic relations, case marking, NP animacy, or information structure. Each of these factors has its own ordering preferences and when taken into consideration during the production or evaluation of a sentence, these preferences can work together smoothly or they can clash. This has consequences for the perceived acceptability of a sentence: the more linearization preferences a specific order obeys, the more acceptable it feels to speakers. German grammar does permit a number of possible word orders and therefore a variety of constituents can appear at the beginning of a sentence. Despite this variety, German native speakers nevertheless exhibit a strong preference to interpret the initial argument of a sentence as the subject or agent in ambiguous sentences (see Hopp, 2007; Kempe & MacWhinney, 1999; Schlesewsky, Fanselow, Kliegl, & Krems, 2000). This could be due to the fact that the agent > patient order violates fewer ordering principles than the patient > agent order, or it could be an influence of frequency. Although the German grammar allows flexible word order, Primus (1998) gives an approximate 97% frequency of occurrence for the agent > patient order.

The influence of **thematic relations** and **case marking** on constituent order has been analyzed by Primus (1998) in a general overview of ordering principles in the languages of Europe. She takes the theoretical view of a Generalized Hierarchy Grammar (Primus, 1994) in which every system aligns its relational concepts on a hierarchy including both thematic and formal relations. She decomposes thematic roles that reflect deep case or semantic relations into more basic or prototypical relations. Each thematic role also has a different degree of membership to a prototypical role. Primus lists three prototypical roles: Proto-Agent, Proto-Recipient and Proto-Patient. The Proto-Patient depends thematically on the Proto-Recipient and the Proto-Agent, while the Proto-Agent is thematically independent. The conventional ordering of constituents in ditransitive sentences according to the thematic hierarchy below is: agent > recipient > patient, corresponding roughly to nominative > dative > accusative in German.

Proto-Agent >	Proto-Recipient >	Proto-Patient
Controller Causer Experiencer Possessor	Recipient Addressee Benefactive	Controlled Caused Stimulus Possessed

Table 3.1 Thematic hierarchy of proto-roles as suggested by Primus (1998)

The thematic hierarchy is only one factor that can theoretically influence the ordering of constituents. The second factor is the case hierarchy that orders cases based on their prominence. A universal case hierarchy for nominative-accusative languages can be seen in (24) (in German, only the first four cases are relevant):

(24) nominative > accusative > genitive > dative > instrumental > prepositional

Due to the universal principle of morphological coding, patients are usually coded by a case that is higher on the case hierarchy than the case marking of recipients. The case hierarchy is not sensitive to the thematic hierarchy seen in Table 3.1 and therefore predicts the following preferred order for ditransitive sentences: nominative > accusative > dative. The orders predicted by the two hierarchies obviously clash with regard to the order of Proto-Recipient and Proto-Patient. Primus suggests that this clash is resolved in one direction or the other depending on how informative morphological case marking is. For example, in ditransitive verbs without a formal distinction between Proto-Recipient and Proto-Patient, due to the absence of overt case marking, the basic word order will be predicted by the thematic hierarchy and therefore be recipient > patient. English and Norwegian do not have morphological case marking and should follow this prediction. As can be seen example (25), Primus' prediction is borne out. A reversal of argument order is only possible when the recipient is formally distinguished from the patient by means of a prepositional dative as in (26).

(25) Mannen sender kvinnen brevet.
Man-the send$_{\text{PRES}}$ woman-the letter-the
'The man send the woman the letter.'

(26) Mannen sender brevet til kvinnen.
Man-the send$_{\text{PRES}}$ letter-the to woman-the
'The man sends the letter to the woman.'

In languages like German, and all Slavic languages except Bulgarian and Macedonian, the Proto-Recipient and the Proto-Patient are formally distinguished. The Proto-Recipient receives dative case and the Proto-Patient receives accusative case marking. Dative case is lower on the case hierarchy than accusative case suggesting the preferred patient > recipient order (27a+b) that is at conflict with the recipient > patient order demanded by the thematic hierarchy (28a+b). As the two arguments are formally distinguished, their reversal requires no additional marking with a preposition (although this possibility does exist in German).

(27a) Der Mann schickt den Brief der Frau.
the$_{\text{NOM}}$ man sends the$_{\text{ACC}}$ letter the$_{\text{DAT}}$ woman

(27b) Mužčina posylaet pismo ženščine. (Russian)
man$_{\text{NOM}}$ sends letter$_{\text{ACC}}$ woman$_{\text{DAT}}$
'The man sends the letter to the woman.'

(28a) Der Mann schickt der Frau den Brief.
the$_{\text{NOM}}$ sends the$_{\text{DAT}}$ woman the$_{\text{ACC}}$ letter

(28b) Mužčina posylaet ženščine pismo. (Russian)
man$_{\text{NOM}}$ sends woman$_{\text{DAT}}$ letter$_{\text{ACC}}$
'The man sends the woman the letter.'

Primus claims that it is this conflict between the two hierarchies that causes the free word order in languages like German and Russian, as they can follow either of the two ordering principles. Both resulting argument orders can be considered basic orders according to Primus, even though she takes the thematic hierarchy to have a stronger influence on constituent orders than the case hierarchy. The Slavic languages, for example, do indeed show a slight preference for the recipient > patient order, but this order can easily be reversed to comply with pragmatic principles that will be addressed more thoroughly in Section 5.1.

As could be seen in example (25) and (26) above, Norwegian has lost overt nominal case distinctions for dative and accusative objects. This loss of case marking applies to all of the Mainland Scandinavian languages (Danish, Swedish and Norwegian) and also affects syntactic case distinctions, resulting in a complete absence of case effects on word order. Nominative case is quite rigidly tied to the sentence-initial argument position, while objects are found in a postverbal, VP-internal argument position. Study 1 of this thesis explores a construction – object topicalization – in which this strong tie is broken by having an object in the sentence-initial position. In (29) the object *osten* 'the cheese' has moved from its original postverbal position to the sentence-initial position without changing its syntactic case assignment. The subject *musen* 'the mouse' has remained in its original position and receives subject case there. Due to the absence of morphological case marking the movement of the object is not immediately recognizable for the parser. It will likely propose a subject-initial sentence based on the higher frequency of subject-initial compared to object-initial sentences in Norwegian, but not based on syntactic case assigned in SpecCP as it would be the case in English (see schematic in Figure 2.1).

(29) Osten vil musen spise.
cheese.the will$_{\text{PRES}}$ mouse.the eat
'The cheese, the mouse will eat.'

A diminishing relevance of case leads to an increasing loss of overt case marking in the Germanic languages, but according to Primus, the size of the remaining case system allows no inferences with regard to the syntactic relevance of case. The topicalization of a non-agent for example follows similar syntactic rules in Norwegian and Icelandic, although the former has lost most of its case marking, while the latter language has retained a rich articulated case system.

Case was manipulated differently in Studies 1 and 2. Due to the absence of overt nominal case marking in Norwegian, the distinction between subject and object in Study 1 is achieved through the use of a word order that unambiguously signals the status of each argument. Study 2 used argument nouns that had unambiguous case marking to distinguish the subject and the direct and indirect objects. Thematic role assignment in both studies was also signaled by word order and case marking respectively. Additionally, animacy was entangled with thematic role assignment which is relevant especially for the L2 group in case thematic role assignment based on syntactic and morphological cues fails.

The status of **animacy** has been discussed in the theoretical literature and also widely explored in experimental research, as will be seen in the literature review in Section 3.2. In the theoretical literature the debate centers around the question of whether animacy asserts an independent influence on ordering preferences as suggested by Tomlin (1986) in his Animated First Principle which states that animate arguments preferably precede inanimate arguments. Primus (1998) sees no evidence for an independent influence of animacy as it is closely tied to the thematic hierarchy that can solve ordering issues without explicit reference to animacy. Proto-Agents are preferably animate, as are Proto-Recipients. Volitional or sentient agents have to be animate by definition because inanimate entities are not sentient. From this it follows that in a thematically canonical ditransitive clause, animate NPs precede inanimate NPs since Proto-Agent and Proto-Recipient are highly likely animate and the Proto-Patient is inanimate.

In this thesis, animacy was only manipulated in the experiment on Norwegian object topicalization introducing preferred (animate subject) and dispreferred (inanimate subject) associations of animacy and agency. In the German experiment on ditransitive structures all constituents occurred with the expected animacy values. As in example (30) the agent was always animate (*die Enkelin* 'the granddaughter'), the recipient mostly animate (*der Großvater* 'the grandfather'), and the patient always inanimate (*der Kuchen* 'the cake'). Reversing the animacy of the recipient and patient would have resulted in a very limited set of plausible sentences.

(30) Die Enkelin hat dem Großvater den Kuchen mitgebracht.
the$_{\text{NOM/ACC}}$ granddaughter has the$_{\text{DAT}}$ grandfather the$_{\text{ACC}}$ cake brought
'The granddaughter has brought the grandfather cake.'

As already stated above, non-canonical orders like object-first or scrambled sentences are a lot less frequent than subject-first sentences in natural discourse. This is mainly due to the more limited contexts that license object movement. The influence of **information structure** on word order has been studied especially in the Eastern European linguistic schools and the proposed linearization preference is given > new information. I will go into more detail for this tradition in Section 5.1. Canonical word orders are neutral with regard to information structure, and can appear in any context or context-free situations. Sentence (32a) with canonical word order is an acceptable answer to the question in (31). Sentence (32b) with a focused direct object is not an acceptable answer to (31), instead it requires a question asking for a contrastive answer like (33). The scrambled version of (32a) with the accusative > dative order in (34) appears in contexts in which the accusative object is known information as it has been introduced previously and is therefore an answer to a question like (35). (32a) is an acceptable answer also to (33) and (35), because of its wide focus. Taken together, it can be said that objects in non-canonical positions are either given information that has been introduced by the context or they are focused or contrasted information. In the case of focused or contrastive objects, the change in word order also has consequences for the intonation of the sentence as the focused/contrasted object receives contrastive stress.

(31) Was macht der Mann gerade?
what$_{\text{NOM/ACC}}$ makes the$_{\text{NOM}}$ man now
'What is the man doing right now?'

(32a) Der Mann schreibt der Frau einen Brief.
the$_{\text{NOM}}$ man writes the$_{\text{DAT}}$ woman a$_{\text{ACC}}$ letter
'The man writes a letter to the woman.'

(32b) Einen Brief schreibt der Mann der Frau.
a$_{\text{ACC}}$ letter writes the$_{\text{NOM}}$ man the$_{\text{DAT}}$ woman
'A letter, the man writes to the woman.'

(33) Was schreibt der Mann der Frau?
what$_{\text{NOM/ACC}}$ writes the$_{\text{NOM}}$ man the$_{\text{DAT}}$ woman
'What does the man write to the woman?'

(34) Der Mann schreibt den Brief einer Frau.
the$_{\text{NOM}}$ man writes the$_{\text{ACC}}$ letter a$_{\text{DAT}}$ woman
'The man writes the letter to a woman.'

(35) Wem schreibt der Mann (den Brief)?
Who$_{\text{DAT}}$ writes the$_{\text{NOM}}$ man the$_{\text{ACC}}$ letter
'Who does the man write (the letter) to?'

The subject > object order is not automatically the canonical order for all verbs, though. Some verbs differ drastically from the ordering preferences of the majority of transitive or ditransitive verbs, such as experiencer-object verbs, a subclass of psych verbs. They can further be subdivided into dative experiencer-object verbs (36) and accusative experiencer-object verbs (37).

(36) Dem Kind gefallen die Kuscheltiere.
the$_{\text{DAT}}$ child please.3.pl. the$_{\text{NOM}}$ stuffed animals
'The stuffed animals please the child.'

(37) Den Mann interessieren die Nachrichten.
the$_{ACC}$ man concern.3.pl the$_{NOM}$ news.
'The news concerns the man.'

As the names suggest in these verbs the experiencer, a thematic role associated with Primus' Proto-Agent is expressed by either dative case or accusative case. Based on the thematic hierarchy, this means that for these verbs, the object > subject order would be the canonical order.

Verhoeven (2015) used a corpus compiled by the Institute for German Language (IDS) to compare ordering preferences and the role of linearization principles for different types of verbs. Prevalence of object > subject (OS) orders was found to differ depending on verb class. While canonical transitive verbs had a proportion of 3.4% of OS orders, proportions for experiencer-object verbs ranged from 18.03% for accusative experiencer-object agentive verbs to 38.5% for dative experiencer-object verbs. Fronted objects were more frequent with experiencer-object verbs and the effects were stronger in the midfield than with one constituent in the prefield. Case syncretism, especially for accusative case, equal animacy of the lexical NPs and ambiguous verbal agreement as in (38, Verhoeven's 18a), can result in structurally ambiguous clauses. Contextual information is then needed for disambiguation. Verhoeven found no occurrences of OS word order in these cases. The first argument was always interpreted as nominative based on the context.

(38) Und jedes Mal entsetzen die jugendlichen Täter die Richter
and every time appalled the adolescent delinquents$_{NOM/ACC}$ the judges$_{NOM/ACC}$

mit ihrer Kaltschnäuzigkeit – von Schuldbewusstsein keine Spur.
with their coolness of guilt no trace

'And every time the adolescent delinquents appalled the judges with their coolness – no sense of guilt.'

Other findings of Verhoeven's corpus study include differences with regard to pronoun distribution (third person pronouns occurred more frequently as subjects for canonical verbs and as objects for experiencer-object verbs) and animacy (no effect of animacy for dative experiencer-objects, with clear effects of animacy for accusative experiencer-objects). Overall, Verhoeven found clear evidence for dative experiencer-object verbs behaving differently from canonical transitive verbs and also from accusative experiencer-object verbs.

The absence of context that is often found in experimental studies of sentence processing should affect object-first sentences more strongly than subject-first sentences, as it violates the context requirements of fronted objects. Some of the studies reviewed in the following sections have manipulated the presence of context, but results have been mixed. Some found reduced reading times for object-first sentences in a supportive context (Kaiser & Trueswell, 2004; Mak, Vonk, & Schriefers, 2008), others only found improved accuracy scores (Kristensen, Engberg-Pedersen, & Poulsen, 2014) or no influence at all (Slioussar, 2011). The heterogeneity of the findings could be caused by crosslinguistic differences in the influence of context on sentence processing and stresses the different possible areas in which this influence can work – globally (comprehension question response time, accuracy) or locally (reading times at point of disambiguation). Differences in the methodologies and materials could also be responsible for the different findings.

As stated in the beginning of this section, non-canonical word orders have lower acceptability rates than canonical word orders. Generative approaches to word order argue that this is likely to non-canonical word orders being the result of the discontinuous placement of two constituents that belong to the same phrase or of the movement of an entire phrase out of its canonical position. Both operations have consequences for sentence processing and partly also on acceptability ratings (Fanselow & Frisch, 2006). The relationship between the frequency of an order and its acceptability seems to be less clear. Studies comparing acceptability ratings of specific orders to their corpus frequencies have found grammaticality-frequency gaps. Orders with similarly high ratings were found to have very different frequencies, while orders with nearly no occurrence in the corpus received varying ratings (Kempen & Harbusch, 2005, 2008).

A nearly constant finding in reading-time studies of object movement is that non-canonical object-initial sentences give rise to longer reading times, lower accuracy scores and slower response times to comprehension questions (see the literature review in Section 3.2). This is often attributed to non-canonical word orders having a higher syntactic complexity than canonical word orders, or to the fact that non-canonical orders require a reanalysis once the parser reaches the point at which the canonical interpretation has to be abandoned in favor of the non-canonical one.

Chapter 2.3 identified two types of object movement that can be distinguished. One is the movement of an object to the sentence-initial, prefield position that takes placed in the topicalization in (32b) above. The other is the movement of arguments within the midfield – commonly referred to as 'scrambling', as in (34) above. Both of these phenomena were investigated in this thesis. Object topicalization is the topic of Study 1 and scrambling is addressed in Study 2. Depending on the theoretical approach, Study 3 on particle verbs in Norwegian could also be considered as an instance of object movement within in the midfield. Some authors (e.g. Taraldsen, 1983) interpret the alternation in (39a + b) as particle shift, i.e. the object moves past the particle similarly to the more widely discussed object shift in which pronominal objects shift across sentential adverbs (see Section 7.1 for more detail). In this case (39a) would be the canonical order and (39b) the non-canonical order. The opposite view (Collins & Thráinsson, 1996; Hopp, 2007) in which the particle shifts across the object has (39b) as the canonical order.

(39a) David slår på computeren.
David turn$_{\text{PRES}}$ on computer-the.
'David turns on the computer.'

(39b) David slår computeren på.
David turn$_{\text{PRES}}$ computer-the on.
'David turns the computer on.'

In summary, it can be said that non-canonical word orders are special. They require certain contexts in which they can appear, they are less frequent than canonical orders and are perceived as less acceptable. The difficulty they often cause in processing has consequences for a number of factors usually investigated in experimental studies, such as reading times, response times and accuracy. The following literature review presents studies that have investigated a range of structures involving moved objects, garden path phenomena pertaining to subject/object ambiguities and ditransitive word orders. Factors that were investigated include NP animacy, verb type and different types of reanalysis.

3.2 Literature review

3.2.1 Subject/object ambiguities and garden path studies

Research on phrase-structure ambiguity resolution has focused on three main types of ambiguity:

- main verb vs. reduced relative clause ambiguities
- subject vs. object ambiguities
- modifier attachment ambiguities

All three of these types have been investigated more or less extensively in L1 as well as in L2 research. Here, I will focus exclusively on subject vs. object ambiguities as these were investigated in Study 1. Depending on the syntactic properties of the language in question, different structures have been investigated as subject/object ambiguities. The predominant structure in research on English is direct object ambiguities (8 repeated here as 40) whereas in German and Dutch object topicalizations are often investigated (23b repeated here as 41).

(40) Peter knew the answer was false.

(41) ...dass die Frau die Kinder getroffen haben.
that the$_{\text{NOM/ACC}}$ woman the$_{\text{NOM/ACC}}$ children met have
'...that the children have met the woman.'

Frequently employed methods in this line of research are SPR and eye tracking paradigms, but other methods, especially rating tasks, are also popular.

In her review of reading-times studies in L2 ambiguity resolution research Papadopoulou (2005) points out that one main finding in this line of research is a divergence between online and offline tasks, suggesting that L2 competence and processing ability do not develop in parallel. L2 speakers were overall slower than L1 speakers and their parsing decisions and commitments were less robust. The studies reviewed also showed an influence of lexical/thematic cues on parsing decisions in L2 (and L1) and the non-universality of strategies such as Late Closure and Minimal Attachment. Papadopoulou's review reflects the knowledge of the small number of studies that were available ten years ago. It serves as a starting point for the remaining literature review that considers studies published after 2005.

Studies investigating the processing of canonical and non-canonical word orders have also been conducted on non-Indo-European languages that display ordering properties that are not common in Indo-European languages. Erdocia, Laka, Mestres-Missé, & Rodriguez-Fornells (2009) found evidence that in Basque, a head-final language with free word order and an ergative case marking system that was assumed not to have a basic word order, one simple word order is preferred in processing. Even though all permutations of word order are grammatical in Basque, SOV word order was processed the fastest and assumed to be the underlying canonical word order by the authors. In their two self-paced reading experiments and an additional ERP experiment, the authors contrasted SOV (42), OSV word order (43) and fully ambiguous sentences (44). The ambiguity in (44) arises due to the identical case marking (-ak) for singular subjects and plural objects.

(42) Emakume-ek gizon-ak ikusi dituzte.
women-the$_{\text{SUBJ}}$ men-the$_{\text{OBJ}}$ seen have
'The women have seen the men.'

(43) Gizon-ak emakume-ek ikusi dituzte.
men-the$_{\text{OBJ}}$ women-the$_{\text{SUBJ}}$ seen have
'The men, the women have seen.'

(44) Gizon-ak emakume-ak ikusi ditu.
man-the$_{\text{SUBJ/OBJ}}$ woman-the$_{\text{SUBJ/OBJ}}$ seen has
'The man has seen the women./The woman has seen the men.'

They found faster reading times and higher accuracy rates for the SOV conditions. Furthermore, fully ambiguous sentences such as (44) were interpreted as SOV sentences and did not show a significant slowdown in reading times like OSV sentences. An additional finding was that subjects were processed more slowly than objects independent of their position in the sentence, which the authors attributed to the ergative case marking in Basque that requires additional processing efforts on the subject. These experiments support a derivational approach to word order and the presence of a canonical word order that is processed faster than all other orders, even in highly inflected, non-nominative languages.

3.2.1.1 Studies on plausibility and semantic persistence in direct object ambiguities

Direct object ambiguities have frequently been used to investigate garden path effects and the influence of additional factors such as plausibility or subcategorization preferences on garden path strength. Studies 1 and 2 of this thesis did not investigate direct object ambiguities themselves. However, other aspects of this research are still relevant to this thesis, particularly the results of plausibility manipulations and the influence of semantic persistence, i.e. the lingering influence of an incorrect parse, are the most relevant.

I will discuss Traxler's (2005) SPR study as an example of a study of direct object ambiguities in L1 that served as a model study for many following studies also in L2. This study manipulated sentence plausibility and verb subcategorization preferences to investigate predictions of garden path and constraint-based models. The design of the study was not a classic 2x2 design and sentences were not presented in a Latin Square fashion; instead a between-subjects design was used. A triplet of sentences formed the base for one experimental item as can be seen in the example sentences (45a-c) below (taken from Traxler 2005:2). The first factor manipulated is ambiguity. Unambiguous sentences contained a comma after the subordinate verb prohibiting the interpretation of the second NP as a direct object. Ambiguous sentences did not contain a comma. The second factor that was manipulated was transitivity. Sentences with an intransitive verb like (45a) can never take a direct object, while the direct object interpretation is possible for the transitive verb in (45b) and (45c). A third factor that was introduced solely for transitive verbs was plausibility of the object interpretation for the NP following the verb. While the direct object interpretation is syntactically possible in (45b) and (45c), only the NP in (45c) could be plausibly interpreted as an object. The direct object interpretation can be abandoned earlier, at the second NP (*table*), for the implausible object in (45b), while it is maintained until the main verb (*stopped*) in (45c).

(45a) When Susan fell(,) the policeman stopped and picked her up.

(45b) When Susan tripped(,) the table crashed to the ground.

(45c) When Susan tripped(,) the policeman stopped and picked her up.

Garden path and constraint-based models make different predictions with regard to the points of reanalysis for these different conditions. In the unambiguous condition both accounts predict that no direct object attachment should take place, as the comma signals the end of the subordinate clause and a new following clause. In the ambiguous conditions without the comma, a garden path account would always predict an initial attachment of the second NP as a direct object. This leads to a syntactically incorrect interpretation for (45a) as *fall* is an intransitive verb, and an implausible interpretation for (45b). Effects of reanalysis should be visible at the second NP in both cases when the wrong interpretation is discovered. In sentence (45c), the second NP is a plausible direct object and this interpretation can be maintained until the main verb. At the main verb all three sentences should show signs of reanalysis. The reanalysis should be harder for more plausible direct objects resulting in higher reading times for (45c) compared to (45b). It should also be harder for verbs with a strong direct object preference causing higher reading times for (45b) and (45c) compared to (45a). Constraint-based models, on the other hand, place more emphasis on the subcategorization preference of the verb. This predicts little to no reanalysis effects for (45a) as the subcategorization preference should bias the reader already towards the correct interpretation. (45b) and (45c) should show longer reading times than (45a), as the subcategorization preference of the verb biases the reader towards an incorrect direct object interpretation at first. Elevated reading times for the plausible direct object in (45c) compared to the implausible direct object in (45b) are also predicted by a constraint-based model. The main difference between the garden path and the constraint-based model is therefore in the prediction for sentences of the (45a) type. Accuracy data of the comprehension questions showed overall high accuracy scores of 94% to 98% with no effect of the experimental manipulation. The reading times for the ambiguous conditions without commas showed elevated reading times for implausible post-verbal NPs (45a,b) compared to the unambiguous condition with a comma. Verb subcategorization preferences did not prevent the reader from interpreting the post-verbal NP as a direct object. Elevated reading times on the matrix verb also suggest no influence of subcategorization information on parsing. However, subcategorization preferences did influence the difficulty of processing. Implausible semantic results as in (45a,b) caused more processing difficulty than plausible semantic results as in (45c). The processing of the matrix verb and with it the syntactically disambiguating information was more difficult with increasing direct object preference. One interpretation of this pattern is that subcategorization information is only accessed after the initial attachment needs to be revised supporting a garden path account. Plausibility and subcategorization information cannot prevent misanalyses, but seems to influence how easily a misparse will be abandoned. This study differs from the other studies reported in this chapter as it used a between-subjects design in the SPR task while the others use within-subjects designs. Participants only read one of the three sentence conditions in (45a-c), but saw unambiguous control sentences and temporarily ambiguous sentences. This is probably the biggest weakness of the study as data from different participants was compared across conditions making the results vulnerable to individual differences between participants and strategy effects.

Plausibility has also been found to play a major role in L2 ambiguity processing as learners are influenced more strongly by pragmatic plausibility information than native speakers. Roberts & Felser (2011) conducted an SPR study on the processing of direct object ambiguities by Greek learners of English. The investigators manipulated two factors: the second NP's plausibility as a direct object and the strength of the potential garden path effect. The second NPs in (46a) and (47a) were plausible direct objects, while the second NPs in (46b) and (47b) were implausible direct objects. (46a+b) are complement clauses that should cause weak garden path effects and (47a+b) are preposed adjunct clauses that should cause strong garden path effects.

(46a) The inspector warned the boss would destroy very many lives.

(46b) The inspector warned the crimes would destroy very many lives.

(47a) While the band played the song pleased all the customers.

(47b) While the band played the beer pleased all the customers.

In the L2 group, implausible direct objects showed slower reading times in both garden path conditions from the ambiguous NP on (46b, 47b). In the weak garden path condition, this effect was later reversed as plausible objects (46a) resulted in longer reading times than implausible objects (46b) from the main verb (*destroy*) onwards. This suggests a more effortful reanalysis for plausible objects. The L1 group only showed a small and slightly delayed effect for implausible objects in the weak garden path condition, and no reading time difference in the strong garden path condition. Roberts & Felser suggested that the difference in reading time patterns found for weak and strong garden paths in the L2 reflects a successful reanalysis for weak garden paths, but failure to reanalyze in strong garden path sentences. This finding is in line with the SSH and the greater reliance on non-structural information in L2 speakers. The study by Roberts & Felser was based on an older study by Pickering & Traxler (1998) that had found effects of plausibility in native English speakers using an eye tracking paradigm. However, Roberts & Felser adapted the materials to better suit L2 speakers and changed the method from eye tracking to SPR, they make the point that both changes could be responsible for the absence of plausibility effects in strong garden paths in the L1 group. The authors proposed that due to the overall high reading speed in the L1 group, self-paced reading was not sensitive enough to detect effects of plausibility in strong garden path sentences. A very quick reanalysis at the disambiguating main verb might have been too fast to be reflected in reading time measures. Roberts & Felser also suggest that their changes to the original materials might have erased the plausibility effect in the L1 group. In order to accommodate the L2 group, intervening material between the ambiguous NP and the disambiguating verb was removed. This immediate resolution of the ambiguity could have weakened the plausibility manipulation. However, the Traxler (2005) study reviewed above did find effects of plausibility in L1 processing for structures in which the ambiguity was also resolved immediately and it used an SPR paradigm. Self-paced reading therefore seems able to detect effects of plausibility on the processing of direct object ambiguities, but it is not clear when this effect occurs as two studies using highly similar constructions and the SPR method found different results for L1 processing.

Semantic persistence as found by Christianson, Hollingworth, Halliwell, & Ferreira (2001) was already briefly introduced in Section 2.1.4 within the good-enough processing framework. It describes the continued presence of an incorrect interpretation beyond the point of disambiguation. An eye tracking study by Sturt (2007) investigated the influence of plausibility on semantic persistence. The direct object ambiguities in this study were part of sentence complement ambiguities as in (48a) and (48b). The expectation was that sentence complement ambiguities would be processed more easily than closure ambiguities that had been used by Christianson et al. (2001). Unambiguous control sentences of (48a) and (48b) contained a complementizer *that*.

(48a) The explorers found the South Pole was actually right at their feet.

(48b) The explorers found the South Pole was actually impossible to reach.

Sturt's predictions for the eye tracking data were the following: Based on previous evidence, the phrase *found the South Pole* should be parsed as a transitive VP with a syntactic reanalysis when the next segment *was actually* is encountered. If the initial semantic analysis persists, (48a) then continues in a plausible manner and should show no processing difficulty in later measures. In contrast, (48b) is not a plausible continuation based on the persistence of a transitive VP interpretation and should cause processing difficulty. Sturt found the expected garden path effect on the first-pass reading times of the critical region *was actually*. He also found evidence for semantic persistence indicated by longer reading times in the second-pass reading times for the same region in sentences of the ambiguous (48b) type. No effect of semantic persistence was found for (48a) type sentences.

An eye tracking study by Jacob & Felser (2015) investigated the presence of semantic persistence in L1 and L2 processing of garden path sentences. Participants read garden path sentences containing direct object ambiguities similar to those that had been used in several previous L1 studies (c.f. Christianson et al., 2001). In their materials, Jacob & Felser used two cues to disambiguation: a syntactic cue (the auxiliary *were* following the ambiguous NP *the burgers* in (49)), and a semantic cue (*being reheated*) rendering the previous direct object analysis impossible. To investigate effects of ambiguity length on the magnitude of reanalysis, the length of the ambiguous NP was manipulated by introducing a relative clause. Unambiguous control sentences contained a comma after the subordinate verb of the adjunct clause.

Short [long], ambiguous

(49) While the gentleman was eating the burgers [that were really huge] were still being reheated in the microwave.

Both the L1 and L2 groups showed effects of ambiguity at the point of syntactic disambiguation, but the effects in later measures were stronger in the L1 group compared to the German L2 group. The authors proposed that both groups noticed that the direct object interpretation was incorrect, but the L2 group might not have initiated a full reanalysis. Effects of semantic persistence were also seen in both groups as a general effect of ambiguity at the semantic disambiguation point, and as lower accuracy scores for ambiguous conditions. There was no relationship between L2 proficiency and the effects of ambiguity. Overall, L1 and L2 speakers showed garden path effects and semantic persistence effects suggesting a comparable basic processing architecture in L1 and L2. However, the resolution of the ambiguity and syntactic reanalysis were different in L1 and L2.

Also investigating direct object ambiguities in English with an eye tracking-during-reading paradigm, Hopp (2015) took individual differences between his L2 participants into account when analyzing their ability to use plausibility (50d+e), subcategorization (50b) and morphosyntactic information (case marking) (50c) for reanalysis. The different types of information vary in the strength of their bias towards a direct object interpretation and should correspondingly yield reanalysis effects of different magnitudes.

(50a) When the girl was praying, the boy made some funny noises.

(50b) When the girl was praying the boy made some funny noises.

(50c) When the girl was playing he made some funny noises.

(50d) When the girl was playing the boy made some funny noises.

(50e) When the girl was playing the piano made some funny noises.

Hopp used several measures to calculate individual differences: proficiency, reading speed, working memory, automaticity in lexical processing, and semantic and syntactic integration ability. The latter three measures require some explanation as they are less frequently used in L2 research. Automaticity in lexical processing was assessed by a lexical decision task involving the most frequent words in central semantic fields. Automatization in lexical processing is reflected in low variance in response speed. Semantic and syntactic integration were measured in the same word monitoring task. Participants heard blocks of sentences with normal prose (51a), syntactic prose (51b) that preserved the syntactic context, but exchanged all words with other words from the same lexical category, and random prose (51c) that preserved neither syntactic nor semantic context. Participants saw the target word *diamond* on the screen and were supposed to press a button as soon as they heard the word in the sentence.

(51a) Her boyfriend gave her a beautiful **diamond** for Christmas.

(51b) The gardener fly him the great **diamond** after mother.

(51c) Power it big rain a she **diamond** flower over.

Syntactic integration ability measured how well participants were able to use the syntactic information in (51b) to predict the upcoming word by comparing reaction times of (51b) and (51c). Semantic integration ability measured the benefit of additional semantic information in (51a) by comparing reaction time of (51a) and (51b).

The L2 speakers generally relied more on plausibility and subcategorization than on case marking in their processing, a result compatible with previous results and the SSH. Only some of the measures of individual differences showed effects on reading times. Automaticity of lexical access had an effect in early measures as L2 participants with higher automaticity scores showed bigger and earlier effects of implausibility and subcategorization than participants with lower automaticity scores. These differences did not persist in later measures. Reading speed showed some interactions with later eye tracking measures for the plausible condition that Hopp compares to the results found by Roberts & Felser (2011). The measure for syntactic integration ability was the one that resulted in most interactions across conditions and reading time measures. A three-way division of the L2 group according to their syntactic integration ability results (low – mid – high) showed different use of plausibility and case marking in the low and high group. The low syntactic integration group showed the longest reading times in the plausible condition and non-native-like use of case marking. The high syntactic integration group, in contrast, showed native-like use of case marking evidenced by short reading times in this condition. Hopp explained the absence of effects of proficiency and working memory in the eye tracking data by the nature of the materials. As the disambiguating information appeared immediately after the ambiguous NP, sentences might not have been difficult enough for differences to surface.

3.2.1.2 Case marking, word order, and semantic information in L1 and L2 processing

The previous subchapter was centered on direct object ambiguities and the influence of plausibility on processing and the persistence of incorrect initial analyses. This subchapter moves away from direct object ambiguities and towards the influence of case and semantic information on the processing of a variety of structures in German and Dutch. Many of the studies on German are based on older work on processing in native speakers by the Bader & Meng group that is reported in Section 3.2.2.

Case marking has been found to be difficult for L2 learners to acquire. Case violations are harder for L2 speakers to detect than agreement violations, and case information seems to be used efficiently in online processing only at later stages in L2 development (Clahsen, Felser, Neubauer, Sato, & Silva, 2010; Hopp, 2006, 2010; Kempe & MacWhinney, 1998; Papadopoulou et al., 2011; Rankin, 2014). The case marking system of a language is also prone to fossilization in end state grammars of L2 speakers (White, 2003). Instead of using case, L2 speakers rely on simpler strategies for agent identification such as a first-noun strategy, or resort to animacy information.

Jackson (2007) conducted a sentence comprehension study with repeated measurements to investigate how the use of word order, case marking and semantic information in the shape of noun animacy develops in intermediate L2 speakers of German. The participants were tested three times over the course of a semester to investigate the development of the ability to use case marking in processing. The study used a 2x2 design with noun animacy (one animate vs. two animate NPs) and word order (subject-first vs. object-first) as factors and all sentences were disambiguated by case marking. In the examples below, the object *Trainer* 'coach' is unambiguously marked for accusative on the article den, while the subjects *Kind* 'child' or *Spiel* 'game' are marked with nominative case.

Subject-first, Animate Subject

(52a) Peter kann sehen, dass das Kind den Trainer ärgert.
Peter can see that the$_{\text{NOM/ACC}}$ child the$_{\text{ACC}}$ coach angers

Object-first, Animate Subject

(52b) Peter kann sehen, dass den Trainer das Kind ärgert.
Peter can see that the$_{ACC}$ coach the$_{NOM/ACC}$ child angers
'Peter can see that the child angers the coach.'

Subject-first, Inanimate Subject

(52c) Peter kann sehen, dass das Spiel den Trainer ärgert.
Peter can see that the$_{NOM/ACC}$ game the$_{ACC}$ coach angers

Object-first, Inanimate Subject

(52d) Peter kann sehen, dass den Trainer das Spiel ärgert.
Peter can see that the$_{ACC}$ coach the$_{NOM/ACC}$ game angers
'Peter can see that the game angers the coach.'

A greater reliance on animacy and world knowledge should make sentences (52c) and (52d) easier as they only contain one animate noun. If the participants applied a subject-first strategy to word order that they also use successfully in their English L1, subject-first sentences (52a and 52c) with canonical word order should show higher accuracy rates than non-canonical object-first sentences. Attention to case marking would result in no differences between the conditions as it is equally unambiguous across all conditions. Jackson found a main effect of animacy as sentences with only one animate entity (52c + 52d) were comprehended significantly more accurately than sentences with two animate entities. Sentences with two animate entities showed an improvement in accuracy over time, from 67% to 80% for subject-first sentences and from 50% to 65% for object-first sentences. Subject-first sentences were interpreted more accurately than their corresponding object-first sentences during the entire testing period. Jackson interpreted these findings as showing that case marking is underused by the L2 speakers during sentence processing. For sentences with only one animate entity she suggests purely lexical processing based on world knowledge - an inanimate entity cannot feel anger - that is independent of the word order manipulation. The presence of two animate nouns requires some form of additional structural processing and the higher comprehension accuracy for subject-first sentences is taken as evidence that the participants used a subject-first strategy instead of using the information given by the case marking. These results support L2 processing theories like the SSH in which L2 speakers have more trouble processing syntactic and structural information than using semantic or real-world knowledge. I see one problem with the animacy manipulation of this study. In the materials section the authors report the conditions to be animate vs. inanimate subject, but the explanation for the effect of animacy is then not based on the animacy status of the subject, but rather on the animacy differential between the two nouns. In the condition with an animate subject, two nouns compete for the role of agent, whereas in the condition with an inanimate subject, only one noun is a plausible agent. The actual role of the animacy of the subject is still not clear. All we know is that it is easier to parse a sentence that contains only one plausible candidate for the agent role than two. In Study 1 of this thesis, I use sentences that contain nouns with different animacy that are both plausible agents. Lexical processing based on world knowledge should be less suitable for these sentences.

In another SPR study, Jackson (2008) investigated the role of case marking and thematic verb placement in the processing of object-first sentences by English L2 German speakers at intermediate and advanced levels. The structure investigated was wh-questions with fronted objects. This self-paced reading experiment also employed a 2x2 design with word order (subject-first vs. object-first) and thematic verb placement (early vs. late) as factors.

Subject-first, Early Verb
(53a) Welche Ingenieurin traf den Chemiker gestern Nachmittag im Café?
Which$_{NOM/ACC}$ engineer met the$_{ACC}$ chemist yesterday afternoon in-the café?
‘Which engineer met the chemist yesterday afternoon in the café?

Object-first, Early Verb
(53b) Welche Ingenieurin traf der Chemiker gestern Nachmittag im Café?
Which$_{NOM/ACC}$ engineer met the$_{NOM}$ chemist yesterday afternoon in-the café?
‘Which engineer did the chemist meet yesterday afternoon in the café?

Subject-First, Late Verb
(53c) Welche Ingenieurin hat den Chemiker gestern Nachmittag getroffen?
Which$_{NOM/ACC}$ engineer has the$_{ACC}$ chemist yesterday afternoon met?
‘Which engineer met the chemist yesterday afternoon in the café?

Object-first, Late Verb
(53d) Welche Ingenieurin hat der Chemiker gestern Nachmittag getroffen?
Which$_{NOM/ACC}$ engineer has the$_{NOM}$ chemist yesterday afternoon met?
‘Which engineer did the chemist meet yesterday afternoon?

In all experimental sentences, the first NP had ambiguous case marking that could either be interpreted as nominative or accusative, while the second NP always had unambiguous case marking. If L2 speakers are able to use case marking information to assign thematic roles, elevated reading times should be found at the second NP whenever it disambiguates to the dispreferred object-first reading (53b and 53d). Manipulating the position of the verb by using two different tenses assesses whether participants assign thematic roles based on the presence of the thematic verb. In the early verb conditions, participants encounter the thematic verb before encountering the disambiguating case information, whereas in the late verb conditions, the thematic verb is in sentence-final position. If thematic role assignment depends on the presence of the thematic verb, processing of object-first sentences should be harder in early verb sentences (53b) than in late verb sentences in which role assignment might be delayed (53d). Jackson found generally higher comprehension rates for subject-first sentences than for object-first sentences in the intermediate (81% vs. 60%), the advanced (87% vs. 78%) and the L1 control group (90% vs. 81%). There was a small effect of verb position on accuracy in the advanced group with a bigger difference in accuracy for the early verb condition than in the late verb condition. In the reading time data, Jackson found that the L1 control group showed effects of processing difficulty for the word order manipulation as object-first sentences were read more slowly at the point of disambiguation, but verb position did not affect the native speakers. She interprets these control findings as supporting the fact the German native speakers assign thematic roles before encountering the thematic verb. Neither of the L2 groups showed this behavior. The intermediate L2 group only showed effects of word order in the sentence-final region suggesting that they had been unable to use the case information provided by the second NP and postponed thematic role assignment until all information was available. The advanced L2 group did show a slowdown at the second NP in the object-first condition suggesting that they could make immediate use of the case information provided. However, this slowdown was influenced by verb position. There was a stronger commitment to a subject-first interpretation in the early verb condition resulting in longer reading times when this interpretation needed to be reanalyzed towards an object-first interpretation. Advanced speakers also showed an effect of verb placement in the sentence-final region. Longer reading times in the late verb condition than in the early verb condition suggest that thematic role assignment was generally delayed until this point in the late verb condition. While more proficient L2 speakers were able to use case marking information in the resolution of subject-object ambiguities, they were still aided by the lexical-semantic information provided by an early appearing thematic verb. Jackson interpreted this as a stronger dependence on lexical-semantic information in L2 speakers despite the ability to correctly exploit morphosyntactic information.

An SPR study by Jackson & Roberts (2010) investigated the role of NP animacy in the online processing of Dutch relative clauses by German native speakers. This study also employed a 2x2 design with the factors of subject animacy (animate vs. inanimate) and relative clause type (subject vs. object). In the examples below the clown is the animate subject and the pies (*taarten*) are the inanimate subject. The point of disambiguation for the type of relative clause is the auxiliary (*heeft* or *hebben*) that agrees in number either with the subject or the object. The auxiliary appears before the main verb (*gegooid* or *geraakt*) that appears in the final position of the relative clause and provides additional information with regard to thematic role assignment.

Subject Relative Clause, Animate Subject

(54a) Voor de kinderen is de clown, die de taarten heeft gegooid, het hoogtepunt van de voorstelling.
for the children is the clown$_{SG}$ that the pies$_{PL}$ has$_{SG}$ thrown the highlight of the performance
'For the children the clown, that threw the pies, was the highlight of the performance.'

Subject Relative Clause, Inanimate Subject

(54b) Voor de kinderen zijn de taarten, die de clown hebben geraakt, het hoogtepunt van de voorstelling.
for the children are the pies$_{PL}$ that the clown$_{SG}$ have$_{PL}$ hit the highlight of the performance
'For the children the pies, that hit the clown, were the highlight of the performance.'

Object Relative Clause, Animate Subject

(54c) Voor de kinderen zijn de taarten, die de clown heeft gegooid, het hoogtepunt van de voorstelling.
for the children are the pies$_{PL}$ that the clown$_{SG}$ has$_{SG}$ thrown the highlight of the performance
'For the children the pies, that the clown threw, were the highlight of the performance.'

Object Relative Clause, Inanimate Subject

(54d) Voor de kinderen is de clown, die de taarten hebben geraakt, het hoogtepunt van de voorstelling.
for the children is the clown$_{SG}$ that the pies$_{PL}$ hit the highlight of the performance
'For the children the clown, that the pies hit, was the highlight of the performance.'

The late position of the main verb also allowed the investigation of L2 speakers' ability to commit to thematic role assignment before the encounter of the lexical verb that provides the final information on the thematic roles of the preceding nouns. Previous studies had suggested that the strength of syntactic role assignment in L2 speakers is influenced by the early presence of the lexical verb and that only L2 speakers at near-native levels were able to show early processing commitments before encountering the lexical verb (Havik, Roberts, van Hout, Schreuder, & Haverkort, 2009; Hopp, 2006; Jackson, 2008). This processing behavior is in contrast to that of native speakers of verb-final languages who have been found to interpret sentences incrementally and assign thematic roles before encountering the lexical verb (Frazier, 1987; Friederici & Frisch, 2000; Konieczny, Hemforth, Scheepers, & Strube, 1997; Schriefers, Friederici, & Kühn, 1995). In their offline acceptability rating task, the authors found the same pattern in the L1 and the L2 groups, with better ratings for subject relative clauses than for object relative clauses and better ratings for animate than inanimate subjects. Overall, object relative clauses with inanimate subjects received the worst ratings from both groups. The online reading-times also showed a similar pattern for both groups with the longest reading times for object relative clauses with inanimate subjects and no difference in reading times for both types of relative-clauses with animate subjects. However, the timing of this effect was different: it appeared earlier for the L2 speakers, at the point of disambiguation, whereas for L1 speakers it was delayed until the following segment. The authors interpret the absence of a reanalysis effect as a delay in the assignment of grammatical roles in the face of conflicting information from noun animacy and topicality. The L2 group also showed a general preference for animate antecedent nouns over inanimate antecedent nouns reflected in faster reading times for the former group, while this preference was absent in the L1 group. As a preference for animate antecedents had been found in L1 Dutch speakers in other studies, the authors explained the lack of this effect in their study by the L2 group's greater sensitivity to semantic information and overall high reading speed in the L1 group, which made it impossible to detect signs of parsing difficulty earlier in the sentence. In this study, L2 speakers performed like native speakers as both groups showed signs of reanalysis in the form of longer reading times for the object relative clauses. The L2 group showed this effect directly at the point of disambiguation, the auxiliary, suggesting that their thematic role assignment happened incrementally, and was not delayed until the following main verb. However, this study has the same problem with the animacy manipulation as the Jackson (2007) study reported above, leaving it unclear whether the results are due to the influence of animacy itself or the numbers of animate NPs used in the sentences.

Stoops, Luke, & Christianson (2014) investigated the interplay of animacy and morphosyntactic information in the processing of non-canonical SOV-and-V structures in Russian. In their SPR study, they manipulated the animacy of the arguments and the informativeness of the case markings. Sentences were either globally ambiguous (55a), or locally ambiguous with two different types of local ambiguity. (55b) contained a syntactic local ambiguity as both nouns were marked with the same uninformative case marking, but only one was a potential actor. (55c) contained a semantic local ambiguity as both nouns were animate and potential actors, but the first noun was unambiguously marked as nominative. Finally, there was an unambiguous control condition (55d) in which the first noun was unambiguously marked as nominative and was the only potential agent in the sentence. The participants answered a comprehension question after each trial assessing their interpretation of the second verb (55e). Examples are taken from Stoops et al. (2014:586):

(55a) Rys' lan' počuyala i nastorožilas'.
S O V and V
Bobcat$_{\text{NOM/ACC}}$ fallow deer$_{\text{NOM/ACC}}$ sensed and pricked up its ears.
'The bobcat sensed a fallow deer and pricked up its ears.'

(55b) Rys' von' počuyala i nastorožilas'.
Bobcat$_{\text{NOM/ACC}}$ bad smell$_{\text{NOM/ACC}}$ sensed and pricked up its ears.
'The bobcat sensed a bad smell and pricked up its ears.'

(55c) Lisa lan' počuyala i nastorožilas'.
Fox$_{\text{NOM}}$ fallow deer$_{\text{NOM/ACC}}$ sensed and pricked up its ears.
'The fox sensed a fallow deer and pricked up its ears.'

(55d) Lisa von' počuyala i nastorožilas'.
Fox$_{\text{NOM}}$ bad smell$_{\text{NOM/ACC}}$ sensed and pricked up its ears.
'The fox sensed a bad smell and pricked up its ears.'

(55e) Kto nastorožilas'?
'What pricked up its ears?'

Pretesting of the materials had shown that in a sentence completion task, participants co-referenced the second verb with the subject of the main clause in 90% of the cases. Despite high accuracy scores on the comprehension questions of at least 92% per condition, Stoops et al. found an effect of animacy such that accuracy was lower when two animate NPs were present (55a+c). Unambiguous case marking did not have an influence on accuracy. The reading times showed a significant effect of animacy at the conjunction, as sentences with two animate nouns were read significantly more slowly than sentences with only one animate noun. On the second verb the same effect of animacy was found and an additional effect of syntactic informativeness as ambiguous case marking on the first noun (55a+b) slowed down reading times. The same effects of semantic and syntactic ambiguity were also found in the response times for the comprehension questions. The authors have two interpretations for this: either the relevant information was re-processed in order to answer the question, or the parsing process had not yet been completed when the question appeared. Stoops et al. also propose the possibility of different parsing strategies in response to variations in word order. An attentional or strategic shift towards more reliable sources of information could alleviate demands on working memory during processing. This experiment shows that even in a language like Russian with reliable morphosyntactic cues in the form of case marking, morphosyntactic representations can be weak, and sentence processing can be affected by NP animacy.

Building up on previous studies on inflectional morphology as disambiguating cues in L1 and L2 German, Gerth, Otto, Felser, & Nam (2015) compared the processing behavior of three L2 groups from different L1 backgrounds (L1 Italian, L1 Korean and L1 Russian) to investigate possible effects of L1 features on L2 processing. The three L1s had different configurations with regard to the presences of case or verbal agreement. Italian only has verb agreement, Korean has case and Russian has both case and verbal agreement. Participants were at an intermediate to advanced proficiency level and completed an SPR task involving temporarily ambiguous but grammatical object-first sentences. The study employed a 2x2-design with ambiguity status (ambiguous vs. unambiguous) and ambiguity resolution cue (case vs. verbal agreement) as factors. The point of disambiguation was in the same position across all items and conditions, as can be seen in (56a,b) below in which the point of disambiguation is printed in bold. Unambiguous control conditions contained a masculine object that was unambiguously marked with accusative case.

Ambiguous, agreement disambiguation
(56a) Die Prinzessin aus Spanien **haben** die Reiter ganz spontan fotografiert.
the $\text{princess}_{\text{NOM/ACC.SG}}$ from Spain have_{PL} the horsemen quite spontaneously photographed.
'The horsemen have photographed the princess from Spain quite spontaneously.'

Ambiguous, case disambiguation
(56b) Die Prinzessin hat plötzlich **der** Reiter ganz spontan fotografiert.
the $\text{princess}_{\text{NOM/ACC.SG}}$ has suddenly the $\text{horsemen}_{\text{NOM}}$ quite spontaneously photographed.
'The horseman has suddenly photographed the princess quite spontaneously.'

Gerth et al. found generally lower accuracy scores for ambiguous sentences compared to unambiguous sentences across all four participant groups. The L1 Russian group did not differ from the control group and the L1 Korean group showed greater difficult with ambiguous sentences in the verbal agreement condition. Only a general effect of proficiency was found reflecting higher accuracy scores with increasing proficiency. In the analysis of the reading times, effects of ambiguity were found solely in the spillover region for the agreement conditions, and in the disambiguating region and the spillover region in the case conditions. No effect of L1 background or proficiency was found in the verbal agreement conditions. In the case conditions there was an effect of proficiency in the L1 Russian group, reflecting faster reading times with increasing proficiency independent of the manipulation. Two smaller deviations from the overall similar pattern across groups were found. The L1 Korean group showed an early effect in the agreement condition, while the L1 Russian group showed a delayed effect in the case condition. Overall, Gerth et al. interpreted the results as suggesting that all L2 groups were sensitive to the case and verbal agreement manipulation even at a lower proficiency level than previously tested, regardless of L1 background.

3.2.1.3 Studies on Scandinavian languages as L1 or L2

Processing studies on Scandinavian languages are rare, especially on the topic of object topicalization. To my knowledge, there is only one published study on Danish (Kristensen et al., 2014) that is similar to my study reported as Experiment 1b in Section 4.4. Like the study by Gerth et al. (2015) and my own study, it investigated comprehension of object-initial sentences using an SPR task. Its main aim was to investigate effects of context on the processing and comprehension of OVS sentences by native speakers. It did therefore not include an L2 group, nor did it manipulate any additional cues within the experimental sentences other than the SVO and OVS word order. Kristensen et al. argue that the main reason object-initial sentences persist in language despite causing major processing difficulties, and despite often being misinterpreted, is that in actual discourse context provides enough information to lower the processing costs and to lead to the correct interpretation. This context is often not provided in linguistic experiments which according to the authors affects object-initial sentences more negatively than subject-initial ones. The Kristensen et al. study therefore contrasted the comprehension and processing of object-initial (OI) sentences with that of subject-initial (SI) sentences in contexts that were either neutral, or supported the information structure of the following experimental sentence. The measurements taken in this study were reading times, response accuracy and response speed. If context lowers the processing difficulty of OIs, reading times and response speed should decrease after a supportive context, while response accuracy should rise. The authors also propose that due to the absence of case markings on nouns in Danish, sentence-initial full NPs are ambiguous between subject and object, so the role of word order and contextual information will be even more important in Danish than in other previously tested languages. The study employed a 2x2 design with the factors word order (SVO, OVS) and context (supportive, neutral). As can be seen in Table 3.2, the neutral context was the same for both word orders, while the supportive context naturally changed depending on the word order. The authors used sentences containing non-finite verbs, as the placement of the non-finite verb (*invitere* 'to invite') and a sentential adverb (*dog* 'however') disambiguates between an SVO (57a) and an OVS word order (57b). Examples adapted from Kristensen et al. (2014:130) with the region that was analyzed printed in bold.

(57a) Peter ville **dog invitere Anne** til festen.
Peter would however invite Anne to party-the.
'Peter would, however, invite Anne to the party.'

(57b) Anne ville **Peter dog invitere** til festen.
Anne would Peter however invite to party-the
'Anne, Peter would, however invite to the party.'

As can be seen from the examples, if main verb and sentential adverbial are placed before an NP, this NP is the object of the sentence. If they are placed after an NP, this NP is the subject of the sentence. (The same contrast also applies in Norwegian and will be used in Study 1 reported in Chapter 4.)

The context consisted of two sentences: the first sentence explicitly mentioned the topic of the target sentence, while the second context sentence contrasted the topic with a possible set of candidates, and introduced the second NP of the target sentence. In order to keep the same structure of two context sentences, neutral contexts contained Xs instead of words with the aim to minimize their informational content. This masking with Xs might have been slightly unnatural, but the authors argue that it should not affect the results as it applied to both word orders in the same way.

<table>
<tr><th></th><th>Context 1</th><th>Context 2</th><th>Target</th></tr>
<tr><td>Supportive context
Subject-initial</td><td>Denne historien handler om Peter.
'This story is about Peter.'</td><td>De andre drenge brød sig ikke om Anne.
'The other boys did not like Anne.'</td><td rowspan="2">Peter ville dog invitere Anne til festen.
'Peter would, however, invite Anne to the party.'</td></tr>
<tr><td>Neutral context
Subject-initial</td><td>Denne historie handler om XXX.
'This story is about...'</td><td>XXX XXX XXX XXX XXX</td></tr>
<tr><td>Supportive context
Object-initial</td><td>Denne historie handler om Anne.
'This story is about Anne.'</td><td>Peter brød sig ikke om de andre piger.
'Peter did not like the other girls.'</td><td rowspan="2">Anne ville Peter dog invitere til festen.
'Anne, Peter would, however, invite for the party.'</td></tr>
<tr><td>Neutral context
Object-initial</td><td>Denne historie handler om XXX.</td><td>XXX XXX XXX XXX XXX</td></tr>
</table>

Table 3.2 Context design in the study by Kristensen et al. (2014)

The authors found the same overall pattern that had been established in previous studies: longer reading and reaction times and lower accuracy for object-first sentences. With regard to response accuracy, context did not have an overall effect, but it selectively improved the accuracy rate of OVS sentences from 51.1%

in unsupportive contexts to 75% in supportive contexts. The accuracy for SVO sentences remained at 90.6% for both context conditions. The same effect was found for context and answer speed, as a supportive context facilitated reaction times more in the OVS than in the SVO condition. For the reading times, an effect of context was already found in the region before the manipulation for OVS sentences. They were read more slowly after a supportive context than after an unsupportive one which was not the case for SVO sentences. The authors explained this slowdown as the application of an object-initial reading to the first NP. Only main effects of word order and context were found, and there was no interaction of context and word order at the actual point of manipulation. Based on the results of this study, it appears that context does influence the final comprehension of object-initial sentences, but its influence on online processing seems to be limited. Kristensen et al. give no explanation as to the mechanisms that underlie OVS sentence processing. They explain their null result for the reading times by the absence of overt morphology to signal semantic role assignment that had been present in previous studies. The speakers in their study had only word order as a cue towards thematic role assignment, which could have been harder to process than case marking morphology. Even though the authors argue that their use of XXXs in the unsupportive context applied equally to both conditions, it introduces an unnatural element into reading and could even raise the participant's awareness of the context manipulation, leading to the adoption of some form of reading strategy. It could have been dealt with by using either no context at all as is the case in many other studies on similar phenomena, or using a truly uninformative context, for example one that introduces the party and the reason for the party in the examples given in Table 3.2.

Production studies are more frequent in research on the Scandinavian languages in L1 and L2 acquisition. Bohnacker (2010) found a prolonged non-target-like production of clause-initial objects in L2 Swedish by speakers who had been immersed in Swedish for less than six years. Their performance suggested an initial transfer of information-structure patterns from their native German that becomes more target-like only after prolonged naturalistic exposure of several years. The L2 group in my Studies 1 and 3 had received considerably less exposure to Norwegian than the L2 group in Bohnacker's Swedish study and could therefore still show signs of L1 transfer.

3.2.2 Processing studies on word order in German embedded sentences

A lot of research has been conducted on the processing of word order in German embedded sentences. I have chosen the results of two research groups as a representative sample, because the structures they investigated were similar to the ones investigated in Study 2 and are the foundation for much of the research on German reported in Section 3.2.1. Both groups focused exclusively on native speakers data and the studies mostly predate the review article by Papadopoulou from 2005. The Bader and Meng group published a range of papers on garden path phenomena in German from the late 1990s to the early 2000s. They used mainly SPR and judgment tasks in their experiments. The Bornkessel and Schlesewsky group started publishing their results in the 2000s, mainly using ERPs to investigate reanalysis processes in German embedded sentences. This chapter aims to highlight their findings on German sentence processing across the two methods.

3.2.2.1 Studies by the Bader & Meng group

This group investigated a number of structures, but the common theme was what they called 'syntactic function ambiguities' (SFA). SFAs were defined as "ambiguities that involve one or more NPs that are ambiguous with respect to their syntactic function" (Bader, 2000:206) and can cause different degrees of garden path effects. The following articles focus only on subject/object ambiguities that are caused by the ambiguous nature of an NP. Ambiguities arising for example from verbs that are ambiguous between a transitive and an intransitive reading are not discussed by the authors.

With regard to case assignment preferences of the parser in cases of ambiguous marking, Bader, Bayer, Hopf, & Meng (1996) proposed two Case Preference Principles. When assigning case to an NP, structural case should be preferred over lexical case that is only assigned as a property of particular verbs, and nominative case should be preferred over accusative case, as an accusative case presupposes a subject, but not vice versa. These Case Preference Principles also explain the reanalysis costs associated with object-first sentences as the parser would always assign nominative case first, and if the case marking is ambiguous this requires a later reanalysis.

Bader & Meng (1999) used a speeded grammaticality judgment paradigm to investigate the effects of syntactic and non-syntactic factors on garden path strength. Altogether, their experiment contained four different embedded sentence structures featuring subject-object ambiguities:

Relative clauses

(58) Maria erzählte mir von der Frau, die$_i$ die Eltern t$_i$ angerufen haben.
Maria told me of the$_{DAT}$ woman, who$_{NOM/ACC}$ the$_{NOM/ACC}$ phoned have
'Maria told me about the woman who the parents phoned.'

Embedded wh-questions

(59) Die Direktorin hat gefragt, welche Lehrerin$_i$ einige der Kollegen t$_i$ angerufen haben.
the direct has asked, which teacher$_{NOM/ACC}$ some$_{NOM/ACC}$ the$_{GEN}$ colleagues phoned have
'The director asked which teacher some of the colleagues phoned.'

Pronoun movement

(60) Die Direktorin hat erzählt, daß sie$_i$ einige der Kollegen t$_i$ angerufen haben.
the director has told that she$_{NOM/ACC}$ some$_{NOM/ACC}$ the$_{GEN}$ colleagues phoned have
'The director said that some of the colleagues phoned her.'

NP movement/scrambling

(61) Die Direktorin hat erzählt, daß die neue Lehrerin$_i$ einige der Kollegen t$_i$ angerufen haben.
the director has told that the$_{NOM/ACC}$ new$_{NOM/ACC}$ teacher some$_{NOM/ACC}$ the$_{GEN}$ colleagues phoned have
'The director said that the new teacher phoned some of the colleagues.'

All experimental sentences had an identical make-up with respect to their NPs. The first NP was always a case-ambiguous singular NP and the second NP a case-ambiguous plural NP. The point of disambiguation was the sentence-final auxiliary that either disambiguated towards an SO order when marked for singular, or towards an OS order when marked for plural. All of the examples above are therefore OS orders. The syntactic structures investigated in this experiment were filler-gap dependencies. Relative clauses and wh-questions like (58) and (59) always contain a filler-gap dependency that links the relative pronoun or wh-element with its gap in the embedded sentence. When changing from an SO interpretation to an OS interpretation only the location of the gap changes, as exemplified by the tree diagrams in Figure 3.1 below. The tree diagram on the left shows a subject relative clause and the tree diagram on the right an object relative clause (diagrams adapted from Bader & Meng, 1999:126).

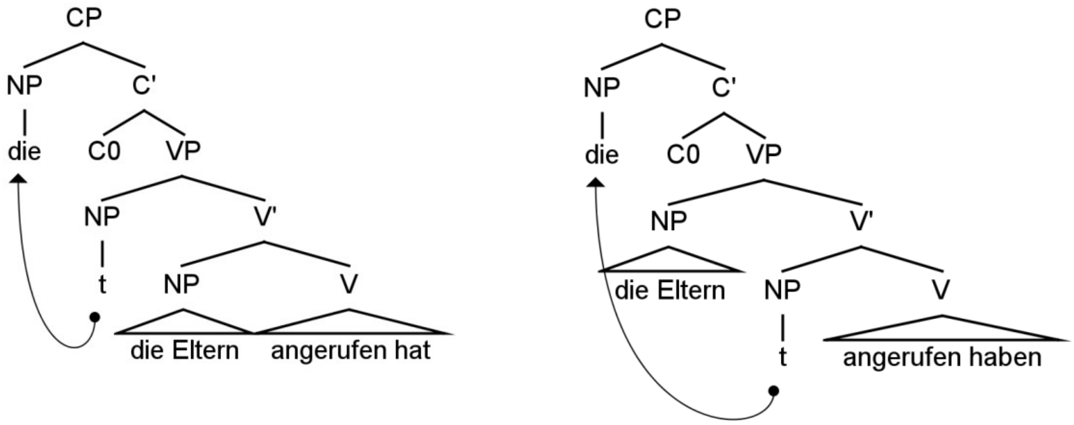

Figure 3.1Tree diagram of subject relative clause (left) and object relative clause (right) in German

In the object relative clause, the gap needs to be posited later than in the subject relative clause. Assuming a first-resort strategy of gap location such as the Active Filler Hypothesis that seeks to identify a gap as early as possible, OS orders should be dispreferred compared to SO orders as they require a reanalysis of the gap location and should lead to garden path effects. According to the authors, pronoun and NP movement as in (60) and (61) differ from (58) and (59) with regard to the presence of a filler-gap dependency in the embedded clause. The SO order does not contain a filler-gap dependency as it can be generated directly from the grammar. The OS order on the other hand contains a filler-gap dependency created by the dislocation of the pronoun or the NP. When the parser has to reanalyze the sentence towards an OS order this filler-gap dependency has to be created, instead of modifying an already existing filler-gap dependency as in (58) and (59). Bader & Meng predicted variations of garden path strength based on the type of syntactic revision needed. If the modification of a filler-gap dependency as in structures (58) and (59) and its creation as in structures (60) and (61) differ with regard to processing effort, there should be differences in judgment accuracy and answer speed for the respective OS structures. The distinction between pronoun and NP movement in (60) and (61) also served to investigate the influence of information structure. Argument order affects the focus/background division in only some of the experimental sentences. A change from an SO to an OS order does not affect the focus division for pronoun movement (60), relative clauses (58) or wh-questions (59) as both orders can be assigned wide focus in these cases. Scrambling on the other hand requires a change from wide focus to narrow focus for its OS order. In (61) repeated here, the subject (*einige der Kollegen* 'some of the colleagues') bears narrow focus, while the moved object NP is backgrounded information. It is therefore not an appropriate answer to the question *What did the director say?*, which requires wide focus.

(61) Die Direktorin hat erzählt, daß die neue Lehrerin [$_F$ einige der Kollegen] angerufen haben.
'The director said that some of the colleagues phoned the new teacher.'

If this additional change in information structure affects garden path strength in any way, an additional difference between the OS orders in the pronoun movement condition (60) and the NP movement condition (61) should emerge.
The accuracy data from German native speakers revealed a garden path effect in the OS orders for all four constructions investigated. SO orders were judged more accurately than OS orders overall (85% vs. 49%) and in each construction separately. The accuracy scores for OS orders in the scrambling condition (61) were also significantly worse (34%) compared to those of the other three conditions (53-55%). Bader & Meng also found differences between the conditions in the SO orders, as those SO orders that involve a filler-gap dependency (58+59) were judged less accurately than those without a filler-gap dependency (60+61) (79% vs. 90%). In order to estimate garden path strength, the authors examined the difference in accuracy scores between the SO order and the OS order. This measure showed a three-way division in which the scrambling condition showed the strongest garden path effect followed by the pronoun movement condition, while the two conditions involving filler-gap dependencies (relative clauses and wh-questions) were similar with regard to garden path strength. The reaction time data also showed a trend towards overall slower reaction times for OS orders, but could not be analyzed more thoroughly due to empty cells, i.e. participants had either misjudged all items in one condition or items had been misjudged by all participants in one condition. Overall, the judgment data showed the predicted preference for SO orders and garden path strength also varied depending on the structure investigated. Establishing a filler-gap dependency is inherently costly: SO orders that contained a filler-gap dependency (relative clauses and wh-questions) had lower accuracy scores than

SO orders without filler-gap dependencies (pronoun and NP movement conditions). The accuracy scores of the OS orders did not show an effect of condition. Modifying an already existing filler-gap dependency was just as costly as creating a new filler-gap dependency. The additional change of focus structure necessary in OS orders in the NP movement condition led to the strongest garden path effect. Bader & Meng interpret this as evidence for an independent contribution of syntactic and non-syntactic factors to the disadvantage of OS orders in the processing of NPs that are ambiguous between subjects and objects. They also suggest that the non-syntactic factor of focus structure might even have a stronger influence than syntactic structure as the difference between the two types of syntactic reanalysis (modification vs. creation of filler-gap dependency) was very weak.

This study is problematic in several different ways. The variation of structures investigated and the alternation of modification and creation of the filler-gap dependency allows interesting comparisons, but the design has some flaws. The influence of focus structure was not manipulated independently from the word order manipulation, and was in fact a confound in the NP movement condition rather than a factor in its own right. Independent claims with regard to the influence of non-syntactic information are questionable, if this "non-syntactic" information is encoded by syntactic movement operations. The two conditions with the relative clause and the wh-question have the same make-up with regard to the presence of a filler-gap dependency and the focus structure, and the authors make no predictions that they should show different effects. They could be reduced to one condition and still have the same explanatory value as the two conditions. The fact that the authors discarded large portions of the data suggests that the experiment might have been inherently problematic for the participants: there were items in which not a single native speaker gave a correct answer to the grammaticality judgment. Given the relatively high number of 56 participants, this is a great cause for concern.

Apart from trying to extend then-current models of serial sentence processing, Bader (2000) also reviewed experimental evidence on the processing of two kinds of subject-object ambiguities in German, both occurring in verb-final structures. The first is an ambiguous filler-gap dependency involving either a relative pronoun (62a) or a wh-pronoun. The scheme of this dependency is given in (62b) (both examples taken from Bader 2000:206). In this type of SFA, the NP1 has been moved to the specifier position of CP leaving a trace, but ambiguous case marking leaves the relative clause locally ambiguous between a subject relative clause and an object relative clause until the encounter of the auxiliary.

(62a) Das ist die Frau, die die Mädchen gesehen hat/haben.
'This is the woman who has seen the girls/who the girls have seen.'

(62b) [$_{CP}$ NP1$_j$ [$_{C'}$ … (t_j) … NP2 … (t_j) … verb(s)]]

The second structure involves embedded verb-final clauses introduced by a complementizer (examples 63a and 63b, scheme 63c). The ambiguity here lies in the possibility to initially interpret the sentences as either an SO or an OS order as case markings are often ambiguous and word order is free.

(63a) …daß der Lehrer dem Schüler zugehört hat. Subj > Obj

(63b) …daß dem Schüler der Lehrer zugehört hat. Obj > Subj
'that the teacher listened to the pupil.'

(63c) [$_{CP}$ [$_{C'}$ daß … NP1 … NP2 … verb(s)]]

Due to a general subject-first preference, object-first sentences were harder to process and showed garden path effects at the point of disambiguation. For ambiguous case marking, Bader proposes that once case has been assigned to an NP, the parser ceases to consider that other case assignments may be possible. When the ambiguity is resolved and a case mismatch arises, this information needs to be recovered through renewed access to the lexicon. The ease of this reaccess is determined by the time that has passed since the encounter of the misinterpreted word. Bader also argues that not all object-first sentences lead to equally strong garden path effects. Object-first sentences exhibiting a base-generated order, e.g. through the use of psych-verbs or ergative verbs, lead to weak garden path effects only causing longer reaction times, but no drop in accuracy for grammaticality judgment tasks. Object-first sentences with a derived word order should cause longer reaction times and a sharp drop in accuracy exhibiting strong garden path effects. This article seems like a review in that it reports several studies and findings, but it does not report any details with regard to participant numbers or more detailed statistics, leaving the reader to trust in the author's interpretation of the experimental results. The structures reported are very similar to the ones investigated by Bader & Meng (1999) and could be from a replication experiment, but without additional elaboration the origins of the results in Bader (2000) remain unclear.

Meng & Bader (2000a) compared case and verb agreement disambiguation and their consequences on garden path strength. Garden path effects for object-first sentences were present in both conditions, but agreement disambiguation resulted in stronger garden path effects than case disambiguation, which was interpreted as reflecting a greater difficulty in reanalysis for agreement disambiguation. The Mismatch Effect, coined by Meng & Bader (2000b), defines the strength of a garden path effect as depending on the salience of the temporary ungrammaticality. If the ungrammaticality at the point of disambiguation is very salient, the reader will reject the sentence as ungrammatical rather than attempt a reanalysis. If the ungrammaticality is less salient, reanalysis will be chosen. In sentences like (64a) the ungrammaticality is poorly detected (61% accuracy), while in sentences like (64b) the ungrammaticality is detected with nearly 97% accuracy.

(64a) *...daß die $_{\text{NOM/ACC}}$ Frau selten eine Postkarte geschickt wurde.
...that the woman seldom a postcard sent was.

(64b) *...daß er$_{\text{NOM.Sg}}$ gestern die Eltern angerufen haben.
...that he yesterday the parents called have.

Saliency in this definition does not solely depend on syntactic properties like base-generated vs. derived filler-gap ambiguities and in both cases, stronger and weaker garden paths can occur. Instead, ungrammaticality detection, and with it the strength of the garden path effect seems to depend more on the type of feature that is involved in creating the garden path. Salient ungrammaticalities contain features that have semantic content and are unambiguously signaled by morphology, e.g. subject-verb agreement. Less salient ungrammaticalities contain features that neither have semantic content, nor are they unambiguously signaled by morphology, e.g. case marking. A violation of verb agreement may therefore be recognized more reliably than case violations.

3.2.2.2 Studies by the Bornkessel & Schlesewsky group

The Bornkessel and Schlesewsky group mainly used ERP data to explore the processing of scrambled structures in German, often using ditransitive sentences similar to the ones used in my Study 2. Other studies from this group focused on the role of canonicity and frequency on processing.

The ERP study by Schlesewsky, Bornkessel, & Frisch (2003) aimed at teasing apart the influence of grammar and working memory on the processing costs of different German word order variations. In order to do this, the authors looked for the presence of the Left Anterior Negativity (LAN), an ERP component that at that time was interpreted as reflecting the additional working memory load associated with displaced arguments. The constructions that were investigated in the experiment were ditransitive main clauses with three possible word orders (NOM – DAT – ACC, DAT – NOM – ACC, ACC – NOM – DAT). The first constituent varied between being a pronominal or an NP.

	NP	pronoun
NOM – DAT – ACC (no scramble)	Gestern hat der Vater dem Sohn den Schnuller gegeben. 'Yesterday the father gave the pacifier to the son.'	Gestern hat er dem Sohn den Schnuller gegeben. 'Yesterday he gave the pacifier to the son.'
DAT – NOM – ACC (scrambled)	Gestern hat dem Sohn der Vater den Schnuller gegeben.	Gestern hat ihm der Vater den Schnuller gegeben.
ACC – NOM – DAT (scrambled)	Gestern hat den Schnuller der Vater dem Sohn gegeben.	Gestern hat ihn der Vater dem Sohn gegeben.

Table 3.3 Experimental conditions in Schlesewsky, Bornkessel & Frisch (2003), canonical word orders shaded

The first row in Table 3.3 reflects a canonical word order in the NP and the pronoun condition. For the second and third row, the NP condition reflects a non-canonical order, while the pronoun in a pre-subject position is the canonical order in German. Assuming a movement account, there is no difference between the NP and the pronoun condition of rows two and three with regard to movement. If the LAN is a reflection of strain on working memory caused by displaced elements, there should be no difference between the NP and the pronoun condition for object-first sentences as the NP and the pronoun have been moved in both conditions. If the LAN instead reflects the cost of processing a non-canonical word order, effects should only be found in the NP condition as the pronoun condition represents a canonical word order. A LAN was found for non-canonical object NPs on the first determiner. No LAN was found for canonical pronominal objects preceding subjects at the same point. The authors interpret this as supporting an interpretation of the LAN signature based on canonicity rather than working memory.

Bornkessel, Schlesewsky, & Friederici (2002) also used ERPs to investigate the role of frequency as opposed to grammatical structure in the processing of fronted objects in German embedded sentences. The study used a 2x2 design with order (subject-first vs. object-first) and object case (dative vs. accusative) as factors. In order to control for frequency differences, embedded sentences were used. In regular transitive sentences such as *Gestern hat der Gärtner den Jäger beruhigt* ('Yesterday the gardener calmed the hunter.'), the order finite verb + nominative/subject was 96 times more frequent than the order finite verb + non-nominative/object in the W-Pub corpus. The verb + non-nominative order also always signals a non-canonical word order. Embedded sentences such as (65a-d) below have two advantages over main clauses. First, subject-initial structures were only eight times more frequent than object-initial structures and there was no difference in frequency between the dative-first and accusative-first sentences. Second, dative-first sentences (65b) offer a canonical interpretation as argument of a passivized verb (65c), while accusative-first sentences (65d) do not allow for a canonical interpretation (examples adapted from Bornkessel et al., 2002).

(65a) ...dass der Jäger dem Gärtner hilft.
...that the$_{\text{NOM}}$ hunter the$_{\text{DAT}}$ gardener helps.
'... that the hunter helps the gardener.'

(65b) ...dass dem Jäger der Gärtner hilft.
...that the$_{\text{DAT}}$ hunter the$_{\text{NOM}}$ gardener helps.
'... that the gardener helps the hunter.'

(65c) ...dass dem Jäger geholfen wird.
...that the $_{\text{DAT}}$ hunter helped is.
'...that the hunter is helped.'

(65d) ...dass den Jäger der Gärtner besucht.
...that the$_{\text{ACC}}$ hunter the$_{\text{NOM}}$ gardener visits.
'...that the gardener visits the hunter.'

As interactionist or constraint-based processing accounts predict more processing difficulty for less frequent orders, the object-initial orders should be more difficult to process than the subject-initial order, but there should not be a difference between the two object-initial orders as their frequency was comparable. A syntax-based processing account predicts no difference between the subject-first and the dative-first order on the first NP as both offer a canonical order interpretation. Reanalysis effects should be found for the dative-first order at the second NP when the non-canonical interpretation becomes obvious. The accusative-first order should cause processing difficulty as it only allows for a non-canonical interpretation. The behavioral data of the experiment only showed a main effect of order as object-initial sentences had higher error rates and longer response latencies than subject-initial sentences. The ERP data showed a broadly distributed negativity for the accusative-first condition in the time window 300-450 ms at the first NP. No such negativity was found for the dative-first condition; instead, an early posterior positivity was found in the time window 300-400 ms at the second NP. The ERP data support a syntax-driven processing account, given that there was no difference between the subject-first and the dative-first order at the first NP, and there was a reanalysis effect for dative-first orders at the second NP. The authors explained the absence of a difference between the two object orders in the behavioral data by effects of frequency on global sentence comprehension.

Reanalysis is often understood as a process affecting and changing the syntactic representation of a sentence. In their ERP study, Bornkessel, Schlesewsky, & Friederici (2003) investigated the possibility of a thematic reanalysis effect in L1 processing of transitive, unambiguously case-marked German verb-final clauses. For the purpose of this study, the authors assume that the thematic structure of a sentence is independent of its syntactic structure. There is no one-to-one mapping of thematic and syntactic functions since, for example, the syntactic role of subject can correspond to the thematic role of agent (66a) or patient (66b) (examples taken from Bornkessel et al. 2003:271).

(66a) John broke a vase.

(66b) John broke a leg.

The absence of a one-to-one mapping also implies that thematic ambiguities can occur independently of syntactic ambiguities. From a syntactic point of view, sentence (66b) does not require a reanalysis, while the thematic role of the subject would need to be reanalyzed from agent to patient when reaching the second NP, assuming that thematic reanalysis exists. The authors additionally assume that thematic parsing is also done incrementally, just like syntactic parsing. Thematic role assignment also occurs independently of the verb as native speakers of verb-final languages have been found to assign thematic roles before encountering the definite thematic information on the lexical verb. As thematic information at this early point is ambiguous, the parser has to follow certain preferences for thematic role assignment. Instead of committing to exact thematic roles, it assigns thematic proto-roles (see Primus, 1998 and Section 4.2.) that subsume several specific thematic roles under one proto-role. The Proto-Agent, for example, includes the thematic roles of agent, causer, experiencer and possessor. The thematic proto-roles also allow the parser to make a first ranking according to hierarchy, as the Proto-Agent is the highest ranking proto-role, the Proto-Patient is the lowest and the Proto-Recipient being in between. In German, native speakers use morphological case marking to establish thematic relations before encountering the lexical verb. Psychological verbs are a special case with regard to thematic role assignment. One of their arguments is attributed the role of experiencer that requires an animate entity able to experience a change in mental state. In subject-experiencer verbs the subject assumes the role of experiencer, whereas in object-experiencer verbs the object assumes this rule. *Fear* and *frighten* are the classic examples expressing the same emotional state as a subject-experiencer verb, and as an object-experiencer verb, respectively. Bornkessel, Schlesewsky, & Friederici contrasted dative active verbs that maintain the normal thematic hierarchy with dative object-experiencer verbs that require a reversal of the thematic hierarchy as the object is higher in the hierarchy than the subject. To create a full 2x2 design, word order was also varied between an SO and an OS order. Examples below were taken from Bornkessel et al. (2003: 279).

SO order, active verb

(67a) ...dass der Priester dem Gärtner folgt.
that the.NOM priest the.DAR gardener follows
'...that the priest follows the gardener.'

OS order, active verb
(67b) ...dass dem Gärtner der Priester folgt.
hat the.DAT gardener the.NOM priest follows
'...that the gardener the priest follows.'

SO order, object-experiencer verb
(67c) ...dass der Priester dem Gärtner imponiert.
that the.NOM priest the.DAT gardener impresses
'...that the gardener impresses the priest.'

OS order, object-experiencer verb
(67d) ...dass dem Gärtner der Priester imponiert.
That the.DAT gardener the.NOM priest impresses
'...that the priest the gardener impresses.'

Based on the results of a previous study (Bornkessel et al., 2002), the authors predicted no difference between the four conditions at the first NP as the parser can predict a canonical passive structure for the OS sentences. Assuming the adoption of this initial interpretation as a passive, a syntactic reanalysis effect is predicted for the OS structures at the second NP when the sentence is disambiguated towards a scrambled structure. The authors made two predictions for the region of the verb. If morphological case plays the most important role in the establishment of the thematic hierarchy, dative object-experiencer verbs (67c,d) should elicit a general effect of thematic reanalysis compared to the conditions with active verbs (67a,b) that do not require a thematic reanalysis. If word order plays an additional role in that the parser assumes that the first argument is always the highest, a different pattern of results is expected. Thematic reanalysis effects should then be visible for dative object-experiencer verbs only in the SO order (67c), and only in the OS order for active verbs (67b). In the behavioral data, object-initial structures showed a higher error rate and longer RTs to the comprehension questions; object-experiencer verbs also showed longer RTs. In the EEG data, the predictions for NP1 and NP2 were borne out, with no difference between the conditions at the NP1 and an early parietal positivity for OS orders at the second NP. At the verb, object-experiencer verbs also showed an early parietal positivity at the 300-600 ms time window. No effect of word order was found at the verb. The authors interpret these findings as evidence for thematic proto-role assignment and creation of a thematic hierarchy *before* the encounter of the lexical verb. This mechanism of thematic role assignment works solely based on morphological case marking and independently of word order and verb information. Thematic reanalysis takes place at the verb, if the previously assumed hierarchy does not match with the information provided by the verb. Bornkessel et al. (2002) propose the following principles for the mechanism of thematic processing:

- If possible, the first argument is assigned the Proto-Agent role.
- Two arguments must always be hierarchically ordered with respect to each other.
- Unless explicitly prohibited, animate and nominative-marked arguments always receive the Proto-Agent role.

The first two principles should apply cross-linguistically, while the third principle might be specific to German. These principles also only apply under the condition that case is unambiguously marked.

Based on the above-mentioned ERP studies, Schlesewsky & Bornkessel (2004) propose two routes for hierarchical argument processing for case marking languages. If morphological case marking is unambiguous, a morphological pathway is used. If morphological case marking is ambiguous, a positional pathway is preferred instead in which the assignment of thematic roles is weaker and only a broad subject/object distinction is made before encountering the lexical verb. This proposal was further supported by an fMRI study on the neuronal implementation of argument hierarchies in German by Bornkessel, Zysset, Friederici, von Cramon, & Schlesewsky (2005), which found activation in two separate brain areas. One showed activation in response to morphological information and its mapping onto the semantic argument hierarchy, and was also sensitive to the informativeness of case marking. Activation in the other area reflected demands in argument linearization. Bornkessel et al. (2005) suggest that this specialization of subcomponents of a bigger neuronal network is likely language-specific.

3.2.3 The application of linear precedence principles in ditransitive sentences

3.2.3.1 Studies on German ditransitive sentences

Linear precedence principles for German and other free word order languages more generally were first proposed by Uszkoreit (1986) within a generalized phrase structure grammar. He suggested that whenever two linear precedence principles conflict, several grammatical orders are possible and that these principles can involve among other things morphological case, thematic and discourse roles and phonological information. Uszkoreit proposed the following five linear precedence principles for German:

- nominative precedes accusative (S > DO)
- nominative precedes dative (S > IO)
- dative precedes accusative (IO > DO)
- pronoun precedes noun (p > n)
- unfocused precedes focused (f- > f+)

These principles have been investigated mainly through corpus-based research, but Rösler et al. (1998) used an ERP paradigm to investigate the influence of violations of linear precedence principles on the processing of German ditransitive verbs. The authors looked at the three linear precedence principles referring to case: S > DO, S > IO and IO > DO. The participants in this study were presented with all of the six possible permutations of ditransitive sentences: from a canonical S-IO-DO order that violates no linear precedence principles (68a), to the DO-IO-S order that violates all three linear precedence principles (68b). In between these two extremes are the other four orders which violated either one or two linear precedence principles.

S-IO-DO order, no violations
(68a) Dann hat der Vater dem Sohn den Schnuller gegeben.
then has the$_{NOM}$ father the$_{DAT}$ son the$_{ACC}$ pacifier given
'Then the father has given the son the pacifier.'

DO-IO-S order, three violations
(68b) Dann hat den Schnuller dem Sohn der Vater gegeben.
then has the$_{ACC}$ pacifier the$_{DAT}$ son the$_{NOM}$ father given

The authors assumed that the IO > DO principle is weaker than the two principles involving the subject and that based on the type and number of violations a ranking of the six word order permutations could be established, showing increasing processing difficulty and decreasing acceptability. The authors used one unusual measure: comprehension time. This was defined as the time participants took until they indicated their readiness to answer the following comprehension question. Comprehension time only showed an effect of subject position: all orders in which the subject was not in the first position showed similar times, and the number of linear principle violations did not play a role. The two orders with the subject in first position differed from each as the order without any linear principle violations had the fastest comprehension times, while the S-DO-IO order (with one violation) was slower than the baseline order, but faster than the remaining four orders. The authors chose the Left Anterior Negativity (LAN) as the ERP component to investigate, as it is associated with increasing demands on processing resources and syntactic violations. First of all, the LAN was only found following the case marked determiners, and not the NPs. A LAN was found on the first determiner if this determiner signaled an object instead of a subject, and on the second determiner if it signaled a direct object instead of an indirect object. The authors interpreted this effect as an extra load on working memory, whenever incoming material cannot be fit into a canonical structure as role assignment needs to be postponed. The results from this study suggest that non-canonical word orders do not only differ from each other with regard to their context suitability, but also with regard to how they are processed. A gradience of acceptability and processing difficulty based on the number and type of linear precedence principle violations could not be established in such a fine-grained manner. The measure of comprehension time is highly unusual and problematic, as it is not clear what participants actually did during this time. They could have finished the processing of the sentence, repeated the sentence to themselves or simply taken a break. The fact that the comprehension time was longer for all conditions in which the subject was not in the first position suggests that there is a difference between subject-first and object-first sentences. The authors give two explanations for this finding. It is possible that the initial processing of object-first sentences was still completed during the comprehension thereby prolonging comprehension time. Alternatively, the initial parse was completed on time, but participants repeated the sentence in order to prepare for the following comprehension question and it was this repetition that took more time for object-initial than subject-initial sentences.

The ability to scramble or topicalize noun phrases and infinitives is a feature that is not unique to German. However, the exact syntactic conditions under which scrambling, topicalization and remnant movement are possible are unique to German and are assumed to pose a challenge to learners of German. Hopp (2005) investigated the German interlanguage of learners with English and Japanese as their native languages to evaluate different proposals related to the acquirability of features in an L2. The different types of transfer and UG access models touched upon in Section 2.2.3 make different predictions regarding whether features that are absent from or realized differently in the L1 can be acquire by L2 learners. Hopp argues that use of statistical frequencies or surface word orders is not helpful in the acquisition of linearization options as non-canonical orders are highly infrequent and also discourse-optional. No Access models that assume the use of general cognitive strategies such as pattern accumulation based on construction frequency and L1 similarity, as well as a low-level analysis of linear surface order would therefore predict a general failure with regard to the acquisition of the structures investigated by Hopp. Partial Access models assume that the L2 acquisition success is dependent on the features already present in the L1. Features of the L2 that are not present in the L1, or that have more extensive feature values than in the L1, would not be completely acquired by L2 speakers. As scrambling is not available in English, native English speakers should fail to acquire scrambled German word orders. Japanese, on the other hand, allows scrambling and topicalization in a highly similar way as German, making direct mapping from the L1 possible. Japanese native speakers should therefore be able to successfully acquire non-canonical German word orders, unlike native speakers of English. Full Access models assume feature acquisition independent of L1 background and predict the successful acquisition of non-canonical German word orders for both English and Japanese native speakers.

In a bimodal (written and auditory presentation) acceptability judgment task participants were presented with scrambling, topicalization and remnant movement of NPs and infinitives. Examples (69a-e) illustrate five of the seven orders tested in the infinitival paradigm. (Remnant topicalization across a scrambled phrase and remnant scrambling across short-scrambled phrase were left out in the examples.) Remnant scrambling across a medium-scrambled phrase (as in 69b) was not tested in the NP paradigm, as the resulting surface word order would be identical with the scrambling of a complete phrase (as in 69a). Note that a high number of movement operations and traces does not necessarily result in an ungrammatical sentence as shown in (69d). (69e) is less complex than (69d), but the movement operations result in an ungrammatical sentence.

	Scrambling of complete phrase
(69a)	Ich glaube, dass [den Wagen zu reparieren]$_1$ Peter schon t$_1$ versucht hat.
	I think that the.ACC car to repair Peter already tried has
	Remnant scrambling across medium-scrambled phrase
(69b)	*Ich glaube, dass [t$_1$ zu reparieren]$_2$ [den Wagen]$_1$ Peter schon t$_2$ versucht hat.
	I think that to repair the.ACC car Peter already tried has
	'I think that Peter already tried to repair the car.'
	Topicalization of complete phrase
(69c)	[Den Wagen zu reparieren]$_1$ hat Peter schon t$_1$ versucht.
	the.ACC car to repair has Peter already tried
	'Peter already tried to repair the car.'
	Remnant topicalization (across scrambled phrase) across finite clause boundary
(69d)	[t$_1$ Zu reparieren]$_2$ glaube ich [t'$_2$ hat Peter [den Wagen]$_1$ schon t$_2$ versucht].
	to repair think I has Peter the.ACC car already tried
	Remnant topicalization across topicalized phrase
(69e)	*[t$_1$ Zu reparieren]$_2$ glaube ich [den Wagen]$_1$ hat Peter schon t$_2$ versucht.
	to repair think I the.ACC car has Peter already tried

All participants irrespective of L1 background or L2 proficiency were able to correctly distinguish between grammatical and ungrammatical movement. An effect of L1 background was only found for infinitival scrambling in that the English L2 speakers were less accepting of it than Japanese L2 speakers. There was no difference between the L2 groups with regard to acceptance of topicalization. Hopp interprets these findings as supporting a Full Access/Full transfer model for the acquisition of German word order variations as the relative discriminations between the orders were the same across L2 groups. The quantitative differences that were found between the L2 groups and the L1 group suggest prolonged difficulties in the correct identification of semantic and information-structural conditions for syntactic reordering. Using L2 speakers with different levels of proficiency from high intermediate to very advanced was a useful way to study a possible developmental trajectory that is suggested by some of the Full Access models, but it resulted in very small participant groups. There were altogether only 13 Japanese native speaker participants divided into two groups and a slightly bigger English group with 26 participants spread across three proficiency levels, so some proficiency groups only had 5 or 6 participants. The number of experimental conditions was also very high, with seven order conditions in the infinitival paradigm and six in the NP paradigm. This meant that in order to see each condition once, 13 trials were needed. Given the small number of participants, this is problematic: it meant, for instance, that there were only 15 data points per condition in the whole Japanese advanced group.

3.2.3.2 Studies on the English dative alternation

Object order in German ditransitive sentences has not been as widely studied in L2 speakers as the dative alternation in L2 English. In both languages, ordering preferences are influenced by a similar set of factors: noun animacy, definiteness, pronominality, information/discourse structure and syntactic weight. I therefore also include research on the dative alternation in L2 English in this review to the extent that results could be comparable to the situation in the German ditransitive structure.

In a study on adult L2 German, Baten & De Cuypere (2014) investigated possible effects of conceptualization transfer in the acquisition of the dative alternation in L2 German. Participants were L1 Dutch speakers and were asked to rate German sentences containing double object constructions (DOCs) (70a) and prepositional dative constructions (PDCs) (70b). The judgment was a 100-split task in which the participants saw pairs of sentences like the examples below and had to split 100 points between the two sentences. This task allows the participants to indicate ambiguity or equal fit (50-50), ungrammaticality (100-0) and also preferences (75-25). There was a Dutch control group that rated translation equivalents of the German sentences in their L1 Dutch. The PDCs were grammatical in Dutch, but their German equivalents were ungrammatical, at least with the verbs used in this study.

DOC-German
(70a) Das Mädchen gibt der Katze Milch.
the$_{\text{NOM}}$ girl gives the$_{\text{DAT}}$ cat milk
'The girl gives the cat milk.'

PDC-German
(70b) Das Mädchen gibt Milch *an die Katze.
the$_{\text{NOM}}$ girl gives milk to the$_{\text{ACC}}$ cat
'The girl gives milk to the cat.'

The authors found that L1 Dutch speakers had acquired a preference for the German DOC over an ungrammatical German PDC. This preference was correlated with a DOC preference for the translation equivalent verb in their Dutch L1. This was interpreted as a sign of conceptualization transfer. However, this study is problematic in several ways. First, there was no German control group, so the authors have no way to know how native speakers of German would rate the same sentences. Any conclusions regarding the L2 group's proficiency (the higher the preference for DOC, the more proficient) based on the assumed behavior of a control group are highly questionable. Second, the PDC is not ungrammatical in German per se, therefore the prediction that L2 learners should rate the DOC as the only possible choice in German is incorrect. It is more restricted and not all ditransitive verbs can have a PDC, but some verbs even have all three possible orders: DAT > ACC, ACC > non-prepositional DAT, ACC > prepositional DAT. As there is no full listing of all verbs that were used in the rating task, it is possible that there were verbs among them that are actually grammatical with PDCs. Third, the German DOC was apparently only presented in the DAT > ACC order and not in the equally grammatical reversed order as "the opposite ordering [...] is grammatically possible, but is very marked and restricted to contrastive contexts" (Baten & De Cuypere, 2014:11). As a consequence, no conclusions regarding the acquisition of German object order permutations in ditransitive sentences are possible from this study.

In a similar task on the English dative alternation, De Cuypere, De Coster & Baten (2014) tested Russian L2 speakers of English. The English dative alternation is a phenomenon of grammatical variation in which the same semantic content is expressed by two different syntactic structures. The dative can either be expressed with a prepositional object (71a) or with an NP object (71b).

(71a) John writes a letter to Mary.

(71b) John writes Mary a letter.

The choice of either structure in L1 is not a categorical one, but rather a gradient phenomenon that is influenced by a number of factors, such as the relative weight of the two constituents, pronominality, definiteness, animacy or accessibility.

Unlike the Dutch L2 group in the previous study, the Russian L2 group in this study showed no preference for either order in English, while they had clear preferences for individual verbs in their Russian L1, ruling out transfer of preferences from the L1. Transfer of ordering preferences could have been possible as Russian and English share some of the semantic and discourse-pragmatic factors that influence object order. De Cuypere et al. suggest that the principles that govern object ordering need to be (re-)acquired in an L2 and are not simply transferred from the L1, even if the principles overlap or are even identical.

Jäschke & Plag (2016) also used a 100-split task to investigate whether German L2 speakers of English are sensitive to linearization principles in the same way as L1 speakers are. They used the materials from the L1 English study by Bresnan & Ford (2010) and also used the data from Bresnan & Ford's L1 groups as control groups. Participants first read a short text that provided some context for the following experimental sentence. They then judged two experimental sentences involving the dative alternation, evaluating their fit into the previous context. The experimental sentences included a variety of factors that are known to influence preferences for either dative structure in native speakers. Overall, Jäschke & Plag found that the L2 group was sensitive to the distribution of the dative alternation in a similar way as the L1 groups from Bresnan & Ford (2010). The L2 group also showed sensitivity to the manipulated factors, but to a much smaller set of factors than the L1 group. Relative syntactic complexity, pronominality of the theme, and definiteness and animacy of the recipient yielded significant effects in the L1 and the L2 group, while pronominality of the recipient, definiteness and number of the theme, and the occurrence of a previous PP in the context only had significant effects in the L1 group. Both groups also showed highly similar effects sizes with recipient animacy having the biggest effect on their ratings, and relative syntactic complexity having the smallest effect. In summary, while the L2 group was aware of the existence of two dative structures, and superficially gave ratings that were similar to the L1 group, the factors that influenced the L2 group's judgments were more limited than in the L1 group, and in one case included a factor (person of recipient) not used by the L1 group at all.

After a comparison of the three studies that used the same method to assess order variations of dative constituents, it is obvious that firm conclusions are difficult. As the studies by Baten and colleagues did not independently assess the proficiency of the participants, their claim that the Dutch L2 group in their study and the German L2 group in the study by Jäschke & Plag show more native-like behavior and a heightened sensitivity to alignment factors with increasing proficiency is difficult to confirm or disconfirm. We simply do not know what proficiency levels are compared across the three studies. Even if ordering principles are not transferred from the L1, they seem to be acquirable. Whether this native-like performance on the task depends on the psychotypological proximity of the L1 and the L2 (the combinations Dutch-German and German-English are closer than Russian-English), the presence of context (no context in the Russian-English study, context in the Dutch-German and the German-English study), the grammaticality of the structure investigated (two grammatical sentences in English, grammatical-ungrammatical opposition in German) or on the overall proficiency of the L2 group is still an open question. De Cuypere et al. (2014) predict that with increasing proficiency, L2 speakers abandon more easily processable structures (the PDC) and the influence of linearization principles (i.e. given-before-new, long-before-short or pronoun-before-noun) increases. If this is correct, choosing a highly-proficient group of L2 speakers should reveal effects also in German word order alternations that are triggered by linearization principles.

The information structure principle given > new is found crosslinguistically, suggesting a possible source of transfer of linearization preferences (e.g. Kaiser & Trueswell, 2004 for Finnish, but see Clifton & Frazier, 2004 for evidence for a more construction-specific application of this principle). In Park's (2011) study on the application of the given > new principle by Korean learners of English, the L2 speakers did not show a consistent application of the principle. In an oral contextualized preference task, the participants heard a context that either introduced a theme or a recipient. Then the participants heard a DOC or a PDC and had to decide which sentence was more appropriate in English. Six dative verbs were used that could take either a DOC or PDC and they were used in both the given-theme and the given-recipient condition resulting in

12 experimental trials. In the examples (72) and (73) the given element is printed in bold.

Given theme

(72a) John brought **the pie** to some friends. (given > new)

(72b) John brought some friends **the pie**. (new > given)

Given recipient

(73a) Mary brought some cookies to **the policeman**. (new > given)

(73b) Mary brought **the policeman** some cookies. (given > new)

The task was further split as the given element was either rendered by an NP as in the examples above or by a pronoun. The participants adhered to the given > new principle only for given themes (72a), and for given recipients they either showed a preference for the new > given order when both objects were full NPs (73a) or showed no preference at all, when one object was a pronoun. L2 proficiency had no influence on the application of the given > new principle, nor were there signs of transfer from the Korean L1. Park suggests an implicational scale in the L2 usage of the given > new principle beginning with the appropriate adherence to it with given themes, moving on to given recipient pronouns and finally to given recipient NPs. Park mentions that there were two tasks, the NP task and the pronoun task and it seems that these were conducted separately as the author reports different fillers for each task. It is not clear whether the 23 participants reported either took part in both tasks, or were randomly assigned to either of the two tasks. As there were only six verbs investigated, the number of tokens per condition is also comparatively small, but still higher than in the Hopp (2005) study. The development that Park sees in the application of the given > new principle is not entirely convincing. The preference for (72a) and (73a) could be unrelated to the information structure principle that according to Park is only applied for given themes, but instead reflect a pure preference for a PDC when the two objects are NPs.

Marefat (2005) also found a developmental trajectory of sensitivity to information structure in the dative alternation in Persian learners of English. Learners at an elementary level showed L1 transfer by overapplying the PDC (the only grammatical option in Persian), while high-intermediate and advanced speakers performed like native speakers on an acceptability judgment task. Her elicited production task was more problematic as many learners showed an effect of structural priming in their answers by repeating the question structure and showing no sensitivity to the information structure. She also found no systematic construction preferences for particular verbs in the L2 group that had been present in the L1 control group. Marefat attributes this to differences in the underlying representation that lead to a superficially similar performance, but based on different sources.

Both Park and Marefat claim a development in the sensitivity to the information structural requirements for specific objects orders, and that these (at least for English) have reached a native-like level when speakers reach an advanced level. However, the priming effect found by Marefat makes this conclusion seem questionable.

3.3 Summary

Chapter 3 reviewed the main theoretical approaches to word order and their empirical investigation. Grammatical word order variation is often explained by the application of various linearization principles and the interaction of many factors such as thematic roles, case, animacy or information structure, at the same time. This multitude of factors also affects online processing and acceptability ratings. Canonical orders are processed faster and judged as more acceptable than non-canonical orders, which is mainly attributed to their lower syntactic complexity and overall higher frequency. Experiments on passives or object experiencer verbs have shown that object-initial sentences that are canonical cause less processing disruption than non-canonical object-initial sentences (Bader, 2000; Schlesewsky et al., 2003).

Previous research on locally ambiguous sentences has shown some clear similarities between L1 and L2 processing. L1 and L2 speakers were equally susceptible to semantic persistence. Case marking and verb agreement could both be used for disambiguation (Gerth et al., 2015), although case marking involved more effortful processing and violations seemed to be harder to detect (Meng & Bader, 2000b). Once L2 speakers were able to use case marking during online processing, they were able to assign thematic incrementally before the encounter of a sentence-final main verb in the same way as L1 speakers. Differences between

L1 and L2 processing were mainly found regarding the reliability of processing commitments and the use of non-syntactic information. L2 speakers especially at lower proficiency levels were found to rely on sentence plausibility and verb subcategorization preferences in their processing of local ambiguities, while the role of plausibility remained fuzzy in L1 processing (Hopp, 2015; Roberts & Felser, 2011; Traxler, 2005).

NP animacy was found to be a particularly strong force in ambiguity processing and played a major role in thematic reanalysis. It could overrule morphological case marking in L1 speakers (Stoops et al., 2014) and was a driving force in L2 processing (Jackson, 2007, 2010). However, previous studies using an animacy manipulation always contrasted two animate NPs with an inanimate subject and an animate object. Sentences with two animate subjects were found to be more problematic across the board. In the condition with only one animate NP, the sentence was never reversible and could be disambiguated using plausibility or world knowledge. It is therefore unclear, whether the effect of animacy that was previously found is only an effect of competition between two possible agents compared to only one possible agent or an effect of actual NP animacy. In order to address this question, Study 1 in this thesis uses fully reversible sentences that are also plausible when the roles of inanimate subject and animate object are reversed. As L2 speakers have been found to be sensitive to case marking and verbal agreement in their processing of local ambiguities, Study 1 introduces linear order as a disambiguation cue.

Studies on word order alternations such as German scrambling have shown that L2 speakers can accurately judge sentences involving complex movement operations despite the low frequency of these variations. Studies on the dative alternation have shown a mixed picture, as some L2 speakers made a difference between the two orders, whereas others did not. The general conclusion seems that L2 speakers consider less of the above-mentioned ordering factors when making their decision. L1 research on German scrambling found a gradience in acceptability that reflected an increase of violations of linearization principles. However, this gradience was not found in online processing (Rösler et al., 1998). Previous L2 research focused on major changes to sentence content (subject/object ambiguities) or the general acceptability of order alternations. Little is known about the ability of L2 speakers to perceive and process minor differences in acceptability. Study 2 of this thesis seeks to address these questions by investigating the processing of scrambled sentences that violate a minor linearization principle (IO > DO).

4 Study 1: Object topicalization in Norwegian

The experiments reported in this chapter investigated the role of NP animacy in the online processing of sentences containing topicalized objects. Two approaches that were introduced in Sections 2.1.1 and 2.1.2, respectively, provide the theoretical background: the garden path approach and the Competition Model.

Native and L2 speakers of Norwegian with German as their L1 performed an agent identification task and a self-paced reading task. These experiments expand on previous experiments on processing of object topicalization in two different ways.

Firstly, Norwegian fills a gap in the experimental evidence on subject/object ambiguities: it allows similar object topicalizations as the previously studied languages German and Dutch (e.g. Gerth et al., 2015; Jackson & Roberts, 2010; Jackson, 2007), yet it lacks the case morphology of these languages and is more similar to English in that regard. Instead of case marking or verb agreement that have been used for disambiguation in previous studies on object topicalization, my study uses a subtle change in surface word order as disambiguation cue. Case marking and verb agreement remain ambiguous and completely uninformative. So far, there have been no processing studies on topicalized objects in Norwegian, and only one similar study on Danish (Kristensen et al., 2014). German learners of Norwegian are a good L2 group to compare with the native speakers as their L1 contains additional information (case, verb agreement) that German native speakers have been found to rely on heavily when resolving subject/object ambiguities (Kempe & MacWhinney, 1999; MacWhinney et al., 1984). The L2 group is deprived of these cues in the present study. Additionally, the Norwegian surface word order of object topicalizations is identical to the regular German word order of sentences involving auxiliaries. German native speakers could be more familiar with the Norwegian non-canonical word order that is the only cue provided in this study, but as this order can reflect a canonical and a non-canonical order in German this familiarity is not necessarily an advantage.

The second goal of this study is to systematically investigate the role of animacy in native and non-native online ambiguity processing. While previous studies reported in Section 3.2.1 also investigated animacy, no study systematically varied the animacy of both agent and patient, using fully reversible sentences that could take animate and inanimate subjects and objects. In the self-paced reading experiment, animacy was manipulated as one of two factors along with word order.

This chapter begins with an introduction to the linguistic background of object topicalizations in Norwegian and German (Section 4.1), followed by a pilot study that tested the plausibility of inanimate subjects. In order for the sentences to be fully reversible each sentence had to be plausible with an animate as well as an inanimate subject (Section 4.2). Section 4.3 and 4.4 report the results of the agent identification task and the SPR task, respectively. Finally, chapter 4.5 provides a summary and discussion of the results of this study.

4.1 Background: Object topicalization in Norwegian and German

Norwegian itself has some peculiarities that need to be mentioned in the beginning. There are two officially recognized Norwegian standard written languages – Bokmål and Nynorsk that are both taught in school, but have a distinct geographical distribution. Bokmål is the more widespread written standard used by approximately 85-90% of the population. Spoken Norwegian is largely dominated by the local dialects that are on a continuum between standard written Bokmål and standard written Nynorsk, and can contain features of both. Bokmål also has two gender systems. One is a tripartite system with masculine, feminine and neuter gender, the other is a two-gender system with common and neuter gender. Masculine and common gender are identical in their morphological expression (i.e., both take the definite article *–en*), whereas the feminine gender is increasingly replaced by common gender (Rodina & Westergaard, 2015). Speakers tend to use both gender systems side-by-side using feminine gender for some nouns and common gender for other originally feminine nouns. The variety used in the studies on Norwegian was Bokmål as it is the more widely used variety and the one usually taught to L2 speakers.

In this study, the language under investigation (Norwegian) and the L1 of the L2 group (German) come from two different branches of the Germanic language family. Norwegian represents the North Germanic or Scandinavian branch, while German represents the West Germanic or Continental branch. The languages' common family means that the two languages share many features. For instance, they are both V2 languages: the finite verb is always the second constituent in declarative main clauses. In addition, German and Norwegian pronouns show case marking for subject, object and possessive.

However, there are also important differences between the two languages, as the split between their two sub-branches indicates a difference in syntactic typology: the Scandinavian branch contains only SVO languages, while Continental Germanic is mainly defined as SOV, with an alternation between VO and OV in main and subordinate clauses. Another very obvious difference between the two languages, especially for learners, is the fact that case marking on determiners or nouns and subject-verb agreement are present in German but absent in Norwegian. In Norwegian, case marking on nouns is limited to the s-genitive, while verbs are only marked for tense. Norwegian has a similar case system as English, and an agreement system that is even poorer. Despite a lack of rich case marking and verb agreement, Norwegian word order is not a completely rigid SVO order. The V2 rule actually requires inversion to an XVS structure whenever an adverbial (74a), a subordinate clause (75) or an object (76) appears in sentence-initial position:

Inversion after fronted adverbial

(74a) *Neste uke* møter læreren foreldrene.
Next week meet$_{\text{PRES}}$ teacher-the parents-the.
'Next week, the teacher meets the parents.'

No inversion, adverbial in sentence-final position

(74b) Læreren møter foreldrene neste uke.
'The teacher meets the parents next week.'

Inversion after subordinate clause

(75) *Når læreren møter foreldrene*, snakker de med hverandre.
When teacher-the meet$_{\text{PRES}}$ parents-the, talk$_{\text{PRES}}$ they with each other
'When the teacher meets the parents, they talk to each other.'

Inversion after fronted object

(76) *Bøkene* liker jeg.
Books-the like$_{\text{PRES}}$ I
The books, I like.

A corpus study on closely-related spoken Swedish (Jörgensen, 1976) found a 40% occurrence of the XVS order in declarative sentences summing over all possible fronted elements. Due to the absence of studies on Norwegian, I will also cite studies on Swedish in this chapter. Swedish and Norwegian are very closely related and while the absence of data on Norwegian is regrettable, resorting to Swedish allows us to at least get an idea of what the Norwegian data could look like.

In this study I investigated the role of word order (order of main verb and second NP) and subject animacy in the processing of auxiliary sentences in Norwegian. Sentences with an auxiliary or a modal verb are the only syntactic contexts in which non-pronominal objects can be unambiguously topicalized through derivation from the basic word order, and without adding additional syntactic material like cleft constructions. The change from a subject-first to an object-first sentence is done by superficially switching the order of the main verb and second NP. In a canonical subject-first sentence, the main verb (printed in bold below) follows right after the auxiliary (printed in italics); in a non-canonical object-first sentence, the second noun phrase is placed after the auxiliary. This second noun phrase then has to be interpreted as the subject of the sentence, as only the subject can intervene between the auxiliary and the main verb and has to appear either in the prefield (i.e. the region before the conjugated verb) or the midfield (i.e. the region between the auxiliary and the main verb) (Faarlund, Lie, & Vannebo, 1997: 674).

(77a) Mannen *vil* **kysse** kvinnen. (SVO)
Man-the want$_{\text{PRES}}$ kiss woman-the
'The man wants to kiss the woman.'

(77b) Kvinnen *vil* mannen **kysse**. (OVS)
'The woman, the man wants to kiss.

The syntactic tree representations of (77a+b) in Figure 4.1 and Figure 4.2 follow Åfarli & Eide's (2003) generative analysis of Norwegian auxiliary sentences and object topicalizations. They analyze SpecTP as the usual position of the subject in Norwegian main and embedded sentences. SpecCP is occupied by constituents after topicalization ('tematisering'), and it is not a position the assigns case. Verbs need to move to the head of T' to receive tense.

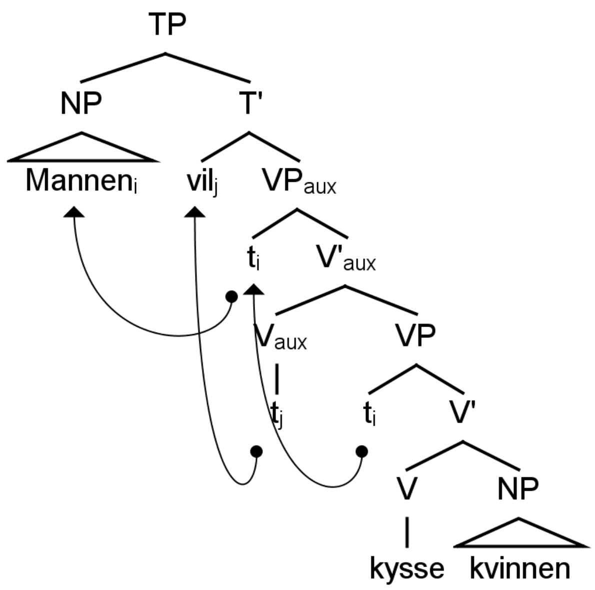

Figure 4.1 X-bar structure of a canonical Norwegian auxiliary sentence

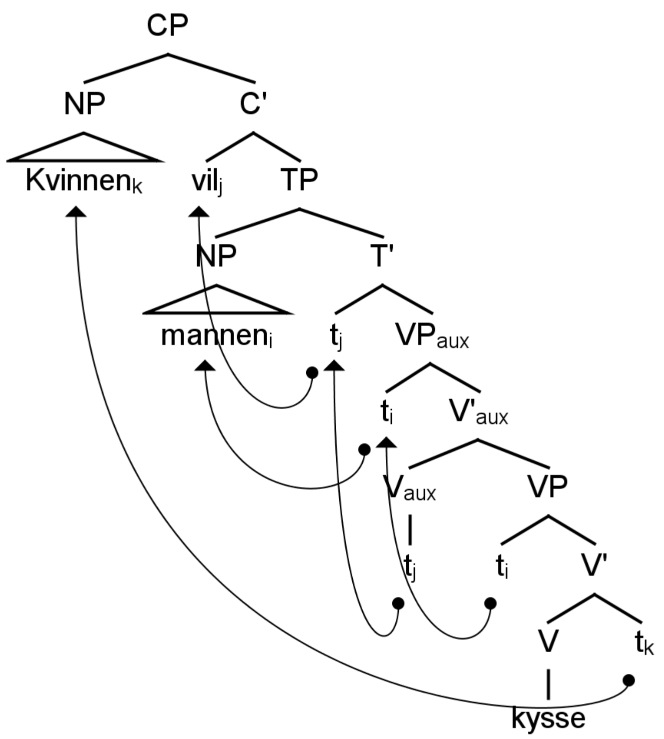

Figure 4.2 X-bar structure of a Norwegian auxiliary sentence involving a topicalized object

As can be seen from the two syntax trees above, the underlying syntactic representation of an OVS sentence differs from its surface representation in more ways than a superficial switch between the main verb and NP2.

The earliest point at which the parser can realize that the current sentence is not an SVO sentence, but instead contains a topicalized object, is when it wants to fill the head of the lower V' with the main verb, but instead encounters an NP (*mannen*). The ensuing reanalysis requires several steps, which I represent here, albeit I do not claim this order to necessarily be the chronological one. As the SVO sentence does not contain a CP, the CP needs to be constructed, and the NP1 (*kvinnen*) has to be moved from SpecTP to SpecCP. During this movement, the NP1 loses its thematic role as agent and its nominative case marking. The auxiliary also needs to be moved from head of T' to head of C', thereby leaving an additional trace at head T'. The second NP that caused the reanalysis needs to be inserted in SpecTP and interpreted as the subject of the sentence. The first VP structure, which contains the base position of the auxiliary, remains unaffected by the reanalysis. The second VP structure can then be filled completely once the main verb is encountered as the next word. The trace left by the NP2 is identified in SpecVP, the main verb is inserted as head of V', and finally the trace left by the NP1 is identified as the complement of V'. Through the establishment of the filler-gap dependency between the NP1 and its trace, the NP1 receives its object case. While the change in superficial word order between the subject-first and the object-first sentence is small and could easily be ignored, it has massive consequences for the syntactic structure of the sentence. Syntax-centric models of processing would predict strong garden path effects created by the number of syntactic revisions needed. Less syntax-centric models such as frequency-based processing models would also predict longer reading times and garden path effects for object-first sentences, as they require an additional thematic analysis (Bornkessel et al., 2003) and are overall less frequent than subject-first sentences.

In sentences without auxiliaries or modals, the order NP1 V NP2 is ambiguous between an SVO and an OVS reading, with the SVO interpretation being the predominant one (Øvrelid, 2004). Without any context, (78) below will most likely be interpreted as an SVO sentence, but with an appropriate context, it could theoretically also be interpreted as an OVS sentence. In the absence of overt case marking on NPs, the surface order is exactly the same for both interpretations.

(78) Foreldrene møter læreren
'The parents meet the teacher.' (SVO interpretation)
'The parents, the teacher meets.' (OVS interpretation)

A corpus study by Øvrelid (2004) on word order interpretations in ambiguous Norwegian sentences of the NP1 V NP2 type (similar to 78) revealed two factors contributing to sentence interpretation in the absence of overt case marking: animacy and definiteness. Subjects were normally higher in animacy than objects with the exception of psych-verbs and causative verbs, which accounted for the majority of sentences with inanimate subjects. Subjects were also usually higher in definiteness than objects, and the indefinite subjects that were present in the sample mainly bore a specific or generic reading. In Øvrelid's sample, only 9.7% of all transitive sentences had OVS order. In this subset, there were no sentences in which the object was higher in animacy than the subject, suggesting that these types of sentences resist topicalization, although the OVS order would not be ungrammatical. Subject and object definiteness did not play a decisive role in OVS sentences, but it was important in sentences with two equally animate arguments as equal animacy of the two arguments also prevented object topicalization.

Cleft sentences such as *Det er foreldrene læreren møter* ('It is the parents the teacher meets.') are frequently used to express topicalization and the relative clause attached to the cleft structure also shows an order alternation between verb and NP to mark a subject or object relative clause similarly to cleft sentence structures in English. Cleft sentences are used at much higher rates in Mainland Scandinavian than in English or German, occurring in 2-3% of all pre-verbal sentence positions in Swedish compared to 0.02% in German (see Bohnacker & Rosén, 2008; Jörgensen, 1976).

German and Scandinavian do differ not only with regards to the occurrence of clefts, but also regarding where new information is placed within the sentence. Production studies by Bohnacker and Rosén (Bohnacker, 2010; Bohnacker & Rosén, 2008) compared the frequencies of various constituents in the German and Swedish prefield and found a stronger tendency to position new information postverbally in Swedish than in German. Swedish also had a higher frequency of subjects in the prefield than German (73% vs. 50%) and lower numbers of objects in the same position (3% vs. 7%). German speakers also fronted more diverse objects than Swedish speakers. The proportions were comparable in written and oral corpora.

Unambiguous object-first sentences in Norwegian are interesting, as the language allows for this word order in very limited syntactic contexts only and has none of the features that are usually manipulated in studies on subject-object ambiguities of other closely-related languages like German or Dutch (Gerth et al., 2015; Jackson & Roberts, 2010; Jackson, 2007). Self-paced reading studies of these languages usually use case or verb agreement for disambiguation. Studies on English, which is more similar to Norwegian in its possibilities for disambiguation, often investigate direct object ambiguities in sentences like *After Bill drank the water proved to be poisoned*. (e.g. Jacob & Felser, 2015; Hopp, 2015; Roberts & Felser, 2011; Sturt, 2007; Traxler, 2005). The present experiment is overall more similar to the previous studies on German and Dutch in that it manipulates the order of the sentence constituents to change its interpretation from a subject-first to an object-first sentence.

All experimental sentences used in Study 1 are syntactic functions ambiguities (cf. Bader, 2000). The Norwegian case system is maximally ambiguous with regard to the case of non-pronominal NPs as there is no morphological case marking either for nominative or for accusative case. Therefore, the first NP is locally ambiguous between a subject or object interpretation until the reader reaches the position after the auxiliary when abstract case is assigned through the order of main verb and the second NP. The two different readings of the ambiguity in the Norwegian sentences investigated in this study do differ prosodically. The SVO order receives a neutral intonation, while the OVS order is usually read with a contrastive intonation stressing the sentence-initial object. Bader & Meng (1999) found that sentences requiring the establishment of a filler-gap dependency, and a change of the sentence's information structure, pose great reanalysis difficulties for the parser. The object-initial sentences investigated in Study 1 should cause major garden path effects as they require both.

Background L2

As previously explained, German and Norwegian are similar to a certain extent, but differ in some crucial aspects such as the use of cleft structures and the amount of non-subject material in the prefield. This difference is what makes German learners of Norwegian an interesting experimental group. Their L1 gives them experience with object-fronted sentences which they may transfer to processing object-fronted sentences in a structurally different L2. For instance, Bohnacker (2010) found that German learners of Swedish produced sentences with more heavy rhematic material and less expletives in the sentence-initial position than Swedish native speakers even after a prolonged exposure. This behavior mirrored their German native language. This could suggest that fronted objects by themselves are not problematic for German learners of Norwegian.

In German the filling of the prefield is obligatory, and the highest element from the midfield - either the subject or a constituent that has been scrambled past the subject - is raised to this position. This purely formal movement does not involve additional semantic or pragmatic features. The prefield can also be filled through base generation of adverbs or movement that causes a contrastive reading of the material in the prefield. The main force behind object fronting in German is information structure together with lexical-semantic factors (Bader & Häussler, 2010).

Figure 4.3 and 4.4 exemplify the syntactic tree structure of the German translations of the Norwegian sentences (77a,b).

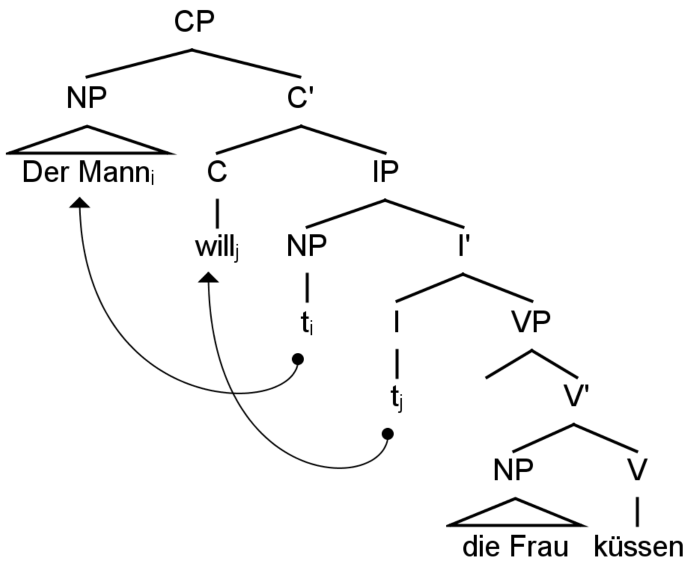

Figure 4.3 X-bar structure of a canonical German auxiliary sentence

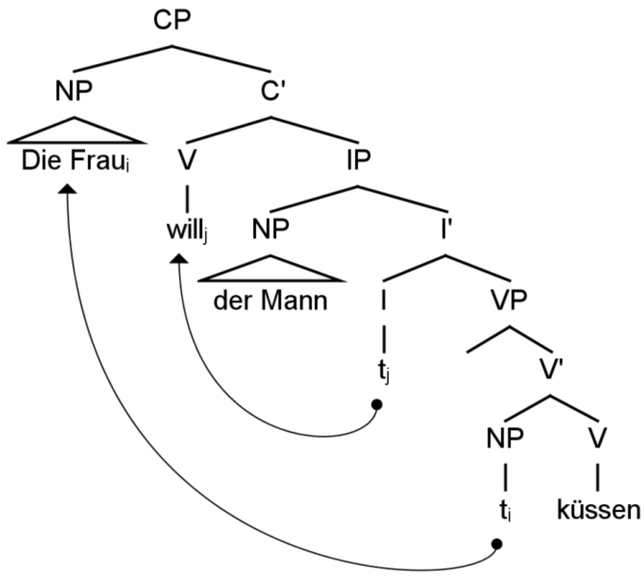

Figure 4.4 X-bar structure of a German auxiliary sentence involving a topicalized object

When comparing the four tree structures with each other, it becomes clear that due to the different order of argument and main verb in the VP in Norwegian and German, the surface word order of the German sentences corresponds to the surface word order of the Norwegian OVS sentence, which might be advantageous for the German learners.

When looking at Study 1 from a Competition Model perspective, it is worth comparing the cues provided by the Norwegian sentences to the cues that MacWhinney, Bates, & Kliegl (1984) suggested as the main cues for agent identification in German native speakers. The Norwegian sentences provide the lexical-semantic cue NP animacy and two word order cues: preverbal, sentence-initial NP position and the order of main verb and second NP. MacWhinney et al. (1984) found the following cue hierarchy for German:

(79) case marking > animacy > agreement > word order

The most important cue for German native speakers, case marking, is completely absent in the Norwegian sentences. Verb agreement that plays a subordinate role in the German cue hierarchy is also absent in Norwegian. Unlike in Norwegian, word order is of little importance in the agent identification process in German. Furthermore, the preverbal, sentence-initial NP position (signaling the subject) is the only word order cue that German shares with Norwegian. The other Norwegian word order cue, the order of main verb and NP2, is not flexible in German: only the order second NP > main verb is grammatical in subordinate and auxiliary sentences. If L2 speakers do not tune their cue hierarchy to the needs of the L2 and continue to use the cue hierarchy of their L1, German learners of Norwegian are deprived of two cues from their L1 and should rely more on animacy than on word order in their assignment of agency as it is the remaining cue that is higher in their native cue hierarchy.

Clahsen & Felser (2006) reviewed evidence that L2 speakers establish filler-gap dependencies based primarily on lexical-semantic and pragmatic information. Similar to the prediction derived from the Competition Model framework, L2 speakers should rely more on the lexical-semantic information provided by NP animacy and possibly plausibility information to assign agency and process the object-first sentences, as opposed to relying on the syntactic information provided by the order of main verb and second NP.

Experiments 1a and 1b of this dissertation investigated the interpretation and processing of object-topicalized Norwegian sentences. In both experiments, the main research objective was to investigate how noun animacy influenced the ease of agent identification (Experiment 1a) and online processing (Experiment

1b). Disambiguation between canonical and non-canonical sentences was signaled by the purely syntactic cue of main verb and NP2 order. The following research questions will be addressed in Experiments 1a and 1b:

Q1.1 Are Norwegian native speakers and advanced German L2 speakers of Norwegian able to correctly interpret sentences with topicalized objects in the absence of a supporting context based only on the order of the main verb and the second NP?

Q1.2 What role do general ordering principles such as animate > inanimate, agent > patient play in the interpretation of object-topicalized sentences? Are sentences harder to interpret if they violate these principles? Is the disruption equally strong for all principles or do some principles have a stronger influence on the interpretation than others?

Q1.3 Is both syntactic (i.e. phrase structure) and lexical-semantic information (i.e. NP animacy) considered during the processing of object-topicalized sentences? Do L1 and L2 speakers differ in their use of the two sources of information?

4.2 Pilot study

Previous research (Pickering & Traxler, 1998; Pickering, Traxler, & Crocker, 2000; Roberts & Felser, 2011; Traxler & Pickering, 1996) has demonstrated that the plausibility of a sentence can influence the ease with which an initial parse will be abandoned during online processing such that plausible parses are more difficult to abandon than implausible ones. Therefore, a preliminary set of sentences created for the SPR task in Experiment 1b was normed for plausibility. One purpose of this pilot study was to ensure that sentences with inanimate subjects were as plausible as those with animate ones. The second purpose was to reduce the set of sentences for the SPR task and choose those sentence pairs – i.e. sentences that are identical except for having subject and object NPs reversed – that have the smallest difference in plausibility. As the test was concerned with the NPs' plausibility as a subject, especially for inanimate NPs, the SVO orders were used in the pretest. Some of the discarded items were later re-used in the agent identification task.

Participants

The participants were 60 native speakers of Norwegian. They were recruited either through personal contact or through posts in online forums, and did not receive a reimbursement for their participation. The software used for this questionnaire recorded the participants' location based on IP address. Five participants had IP addresses outside of Norway (Germany, Canada, United States, UK), but this was not a criterion for exclusion. Data from four participants was excluded as they reported a second native language in addition to Norwegian (Danish, French or English).

The remaining 56 participants (male n=44) had an average age of 28.36 years (range: 17-56, SD: 8.38) and reported no language impairments. Participants also provided information on their use of Norwegian. 51 participants (91%) reported Bokmål as their written variety of Norwegian. The participants spoke dialects from all over Norway: 24 spoke dialects belonging to the Western dialect group (43%), 19 spoke dialects from the Eastern dialect group (34%), 5 participants came from the Trøndersk dialect region (9%) and the remaining 8 participants spoke a Northern Norwegian dialect (14%). 21 reported mixed use of feminine and common gender (37%), 29 participants reported exclusive use of feminine gender (52%) and 6 reported exclusive use of common gender (11%).

Materials

The variety of Norwegian used in this task, and all other experiments on Norwegian, is Standard Bokmål. No feminine gender marking was used, so that all nouns were either marked as neuter or common gender. Common gender was chosen to accommodate the knowledge of the L2 group who might not have acquired feminine gender marking as it is described as optional and not used consistently in the Norwegian text books used in Germany. I wanted to avoid effects based on the feminine gender marking alone such as a slowdown in reading times or an inacceptable rating when the feminine gender is not known to the participant. The two-gender system (common vs. neuter gender) is a grammatical variety that is used predominantly in written Norwegian and expresses a slightly formal style. I expect no negative effects of the two-gender system in the L1 group as they are in daily contact with the two-gender system, even if their local dialect uses a three-gender system. A study by Rodina &

Westergaard (2015) also found that the production of feminine gender in the Tromsø dialect is inconsistent in children (3-13 years old) and young adults (18/19 years) acquiring Norwegian as an L1, while older adults produce it without any difficulty. They interpret this as a sign of grammatical change towards a two-gender system that does not affect the nominal declension system represented by the suffixed definite article. In light of potential problems with the feminine gender for both native and non-native speakers, the two-gender system seemed like the less problematic option.

The pretest employed a 1 (word order, SVO) x 2 (animacy of subject, animate vs. inanimate) mixed design. It was arranged as a Latin Square design so that all participants rated sentences with animate and inanimate subjects. I constructed 37 sets of items with two NPs, one animate, the other inanimate. The verbs I used were transitive verbs that can take both animate and inanimate subjects and objects.

The choice of the tense used in the sentences was limited by the fact that only sentences with auxiliaries or modals are completely unambiguous regarding the assignment of syntactic roles. Simple present and simple past were ruled out as they do not contain an auxiliary, leaving present or past perfect and future tense. Present and past perfect employ the verb *å ha* (to have), while future tense uses the verb *å ville* (to will/to want) as auxiliary. Using either auxiliary could have unwanted consequences on sentence interpretation. The use of present perfect in Norwegian describes actions that are ongoing over a prolonged period of time, therefore this tense is highly problematic with punctual actions like ‘to hit’, as simple past would be the preferred tense in this case.

The verb *å ville* is not only used as an auxiliary in the formation of future tense, but also as a modal verb meaning ‘to want‘. The syntactic surface structure is the same in both cases and the future containing the auxiliary *ville* is the future tense associated with actions on which the subject does not have an influence or that are not planned. The volitional meaning of *ville* is therefore only available in an interpretation as a modal and not as an auxiliary. In (80a) below there is only one possible interpretation of *vil* as an auxiliary for future tense, as inanimate entities can usually not be associated with the action of wanting. As volitional agents always have to be animate, the presence of an animate NP in (80b) allows two readings that are equally possible: *vil* can either be interpreted as an auxiliary for future tense or a modal verb with the meaning ‘to want’ in present tense. This means that (80b) remains globally ambiguous as to whether the action is a volitional, planned one (modal interpretation) or an involuntary one that cannot be influenced (auxiliary interpretation). This asymmetry in the number of possible interpretations of *vil* has to be considered a confounding factor in interpreting the results of the pilot study reported here.

The use of other modals like *kunne* (can) or *matte* (must) was considered problematic with respect to inanimate subjects as well. The fact that the future tense does not require the reader to know participles and does not limit the choice of subjects and/or verbs is taken to outweigh the confound of two possible interpretations of *vil* for animate subjects.

A sample set of items used in the pretest is given in (80a+b). The subject is printed in bold. A full list of all items can be found in Appendix A.

SVO order, inanimate subject
(80a) **Radaren** vil lokalisere soldaten.
radar-the will$_{\mathrm{PRES}}$ localize soldier-the
‘The radar will localize the soldier.’

SVO order, animate subject
(80b) **Soldaten** vil lokalisere radaren.
soldier-the will/want to$_{\mathrm{PRES}}$ localize radar-the
’The soldier will/wants to localize the radar.’

Except for two nouns (*aktivistene* ‘the activists’, *ringene* ‘the rings’), all nouns occurred in the singular. As all nouns had to take animate and inanimate subjects and objects and should ideally be not too infrequent to be known to L2 speakers, the number of possible verbs was small and some verbs were therefore used twice. The nouns were matched according to their length and frequency (see Table 4.1 below). Two sample t-tests comparing length (with and without article) and frequency of the nouns were all non-significant (ps>0.2).

As there was no frequency dictionary available the Oslo corpus of tagged Norwegian texts (http://www.tekstlab.uio.no/norsk/korpus/bokmaal/netscape/treord/oktntb.shtml) was used instead. This corpus contains about 18.5 million words. The frequency search function proved to be of little help as some of the nouns ending in *–er* are homographs of verbs and the frequency search function could not differentiate between those two. As a compromise, the overall number of occurrences in the corpus was used instead. The search function was set to *substantiv* (‘noun’) and *grunnform* (‘basic form’) to search for lemma occurrence. The number of hits was set to the highest number allowed by the system (10.000 hits) and nouns that had more hits than the maximum were replaced in the materials by similar nouns with a lower frequency (e.g. *barn* ‘child’ was replaced by *småbarn* ‘toddler‘)[3].

	animate nouns	**inanimate nouns**
length (without article)	6.05 (1.79)	5.65 (2.04)
length (with suffixed article)	8 (1.93)	7.43 (1.89)
occurrences	803 (1015)	693 (1007)

Table 4.1 Average lengths and number of occurrences of nouns in all 37 sentences, SDs in brackets

The pilot study was conducted to exclude any later effects due to the inanimate subject being completely implausible as a subject. Both types of subjects should at best be equally plausible. The 37 items were spread across two lists using a Latin square design. 12 fillers were added that were either logically impossible or only acceptable if mistakenly interpreted as OVS sentences. The order of items and fillers was pseudorandomized.

Procedure

The lists for the pretest were created using the surveygizmo website (www.surveygizmo.com). Multiple participation from the same IP-address was disabled to avoid that the same participant completed the questionnaire several times and the maximum number of completed responses per list was set to 30. The link to the online questionnaires was distributed through an advertisement on the website www.reddit.com/r/norge and through personal contacts.

The participants completed an online questionnaire asking for a plausibility rating on a Likert scale of 1 to 5. A rating of 1 denoted highly plausible sentences, and a rating of 5 implausible ones. After a short introduction with two examples, each sentence was presented on an individual page. The examples were used to exemplify the two ends of the scale. All ratings were marked as mandatory answers and the participant could not proceed without answering them.

[3] By using this search strategy, the noun *stein* ‘stone’ had 1033 occurrences. Unfortunately, “Stein” is also a masculine name and it was not possible to exclude those occurrences from the results of the regular noun.

At the end of the questionnaire some personal data of the participants was collected to make sure that all the participants were monolingual native speakers of Norwegian and did not have a history of language disorders. Participants were only identified through their IP-address. The completion of the questionnaire took on average less than 10 minutes.

General comment on the data analysis procedure

In the following data analysis and most other analyses in this thesis, I will report between-groups ANOVAs. ANOVAs are run separately on subjects (F_1) and items (F_2) and assume normality or homogeneity of variance. They are used to compare more than two means. They use within-subjects and within-items variables that vary either within the individual subject in the F_1, or within the individual items in the F_2. Between-subjects and between-items variables vary between the subjects or the items. Participant group is a typical between-subjects variable, as a participant can only belong to one participant group, in the case of this thesis, either the L1 or the L2 group. In the item analysis, participant group is a within factor as there are two means for each item, one from the L1 group, the other from the L2 group. The designs used in this thesis are usually within-participant repeated measures designs, so that each participant encounters stimuli from each condition more than once. The factors of the design are therefore within-variables in the ANOVA. A between-groups ANOVA takes the whole dataset consisting of L1 and L2 data and L1/L2 status is entered as a between-variable in the by-subject analysis. The null hypothesis in such tests is that L1 and L2 groups behave identically. Results of an ANOVA are reported as main effects (main effect) that have a general effect on all factors. An main effect of Group in the following analyses would indicate for example that one group has generally faster reading times than the other, independent of the manipulation. Another possible result is an interaction, conventionally indicated by Factor1 x Factor2 that indicates a differential effect of one factor on the other. A Group x Factor interaction in this case means that the effect of factor is either only present in one group or larger in one group than in the other. By convention, an interaction with group is required to be able to split the ANOVA and run separate analyses for each group independently. Reaction time data is usually not normally distributed as it tends to be skewed with the fastest reaction times limited by biological response mechanisms, i.e. there is a limit as to how fast a human can possible respond to a stimulus, but no real upper bound for how slow a response can be. In order to meet the assumptions of the ANOVA data is then transformed to reach normality. I use the box-cox transformation (Box & Cox, 1964) to assess which transformation fits the data best in this thesis. Based on the work with the data from this thesis, a transformation can affect the results in both directions, a significant result can appear, increase or disappear, but it does not have to affect the results.

Results

Across all 37 sentences tested, those with an animate subject received a better rating than those with an inanimate subject, 2.09 (SD: 1.43) and 2.7 (SD: 1.57) respectively. Mean ratings and SDs for each item can be found in Appendix B. The difference between the two ratings was calculated for each item to choose those sentence pairs with the smallest rating difference. I first eliminated all sentence pairs in which at least one sentence had received an average rating of higher than 4.0, leaving 30 sentence pairs. More sentences were eliminated in this way by continually lowering the exclusion criterion until 24 sentence pairs were left. The highest average rating received by any sentence of the remaining 24 pairs received was 3.26, occurring for one sentence from each animacy condition. The biggest difference between the two averages of a sentence pair was an advantage of 1.89 for the animate subject over the inanimate one. The rating difference and the eliminated sentences can also be found in Appendix B.

The final 24 experimental pairs received an average rating of 1.68 for animate subjects (range: 1.04 – 3.26, SD: 0.53) and 2.21 for inanimate subjects (range: 1.1 – 3.26, SD: 0.72). A paired t-test run on the remaining 24 items showed that this difference is significant ($t(23)=-2.96$, $p=0.007$). ANOVAs run on the ratings with gender system, dialect or written language variety as between-subjects factors did not show any significant interactions of any of these factors with the ratings ($ps>0.15$). Length and frequency of the 48 nouns used were still matched after taking out 13 items, all $ts<1$, all $ps>0.65$ (see Table 4.2).

	animate nouns	inanimate nouns
length (without article)	6.13 (1.7)	6.08 (2.14)
length (with suffixed article)	8.08 (1.89)	7.83 (1.97)
occurrences	888.5 (1054)	825.8 (1206)

Table 4.2 Average length and number of occurrences of nouns in the final 24 sentences, SDs in brackets

As expected, all fillers were perceived as implausible by the participants with a mean rating of 4.64 (SD: 0.86). The lowest average rating for a filler item was 3.46, and the lowest average rating on the fillers for a single participant was 3.58. The lowest plausibility rating of any experimental sentence is lower than the most plausible filler sentence that was intentionally designed to be implausible.

Discussion

The attempt to completely match the 24 final experimental items for plausibility was unsuccessful, as the condition with animate subjects was rated as significantly more plausible (1.68) than the condition with inanimate subjects (2.21). This finding is in line with accounts like Primus (1998) that postulate a strong preference for animate Proto-Agents. Unlike in previous studies that manipulated animacy, all sentences can be considered reversible. In the final set of 24 sentences used in the self-paced reading task, no sentence with an inanimate subject and animate object yielded an implausible sentence when reversing the thematic roles as had been the case in previous studies. The average rating of 2.21 for inanimate subjects (out of a maximally implausible 5) suggests that even though animate subjects are better subjects, inanimate entities are not completely discarded as subjects. This is also supported by the fact that some sentences with an inanimate subject even received better ratings than the equivalent sentences with an animate subject. The minimum and maximum average ratings are also comparable, but the inanimate subjects show higher variance and more ratings beyond the 2.0 value. While a perfect matching was not possible, reversible interpretations of the experimental sentences should be possible and a sentence-initial inanimate noun should not immediately be discarded as a potential subject.

Plausibility was not an experimental factor in the SPR task reported in Section 4.4, but given the outcome of the pilot study, it is confounded with the NP animacy manipulation. SVO sentences with animate subjects would be prototypical sentences as they feature the simplest and most frequent word order and the preferred animate subject. The first argument is a perfect Proto-Agent and previous studies showed that the parser tries to assign the role of Proto-Agent to the first argument (Bornkessel et al., 2003). SVO sentences with animate subjects might therefore be processed more easily than SVO sentences with inanimate subjects, although both are canonical sentences. In the OVS condition, sentences with an animate NP1 have the same initial plausibility as SVO sentences with an animate NP1 until the point of disambiguation. As those were rated as more plausible, participants might find it harder to reanalyze OVS sentences with animate NP1s than OVS sentences with inanimate NP1s that have a lower plausibility.

It is not clear whether L1 and L2 speakers will be equally influenced in their processing by this difference in plausibility. Studies on the influence of plausibility in L2 processing (Hopp, 2015; Roberts & Felser, 2011) suggest that L2 speakers rely more on sentence plausibility than on syntactic or morphological cues. The role of plausibility in L1 processing is less clear as some studies found an influence (Pickering & Traxler, 1998; Traxler, 2005), and others found no or only a more limited influence (Roberts & Felser, 2011). Any differences between L1 and L2 speakers in their use of NP animacy will also be evaluated regarding a possible influence of sentence plausibility. It needs to be remembered, however, that the plausibility rating was only run on SVO versions of the sentences, while the self-paced reading task will use SVO and OVS versions.

There was no difference in rating between the different groups based on their use of gender marking, written standard or dialect. These factors should therefore not play a role in the following two experiments, but in order to reduce unwanted confounds, the participants were recruited from the same dialect region.

4.3 Experiment 1A: Agent identification

This task was administered to test whether the participants could make use of the order of main verb and second NP to correctly identify topicalized objects and the agent of a sentence. Detectability of the word order cue was the main focus of this task, in order to identify possible L2 speakers that did not know this cue. NP animacy and plausibility were less important in this task and not systematically manipulated. Filler sentences also checked for other types of non-canonical word orders to assess the more general ability to correctly understand non-canonical word orders.

Participants

65 participants took part in this experiment and were paid either 8 € or 65 NOK for their participation depending on whether they were tested in Germany or Norway. 37 participants were tested in Tromsø, Norway and 28 participants were tested in Potsdam or Berlin, Germany.

The L1 control group consisted of 33 native speakers of Norwegian who were all either students or staff at the University of Tromsø in Northern Norway. All of them had grown up in Northern Norway and spoke that regional variety of Norwegian. The data of one participant was excluded entirely due to extraordinarily long reading times of more than 2 SD above the group mean in the self-paced reading task, and 0% accuracy in one of the conditions in the agent identification task.

The remaining group of 32 L1 participants (male n=9) had an average age of 29.34 years (range: 18-61, SD: 11.12). Four participants were left-handed and one reported being ambidextrous. Three participants reported early exposure to a language other than Norwegian. The languages in question were Spanish (acquired as a second L1 from birth), English (acquired from the age of 2 after moving to the United States) and Korean as a first L1. The participant with Korean as L1 had started to acquire Norwegian as an L1 at the age of four months after an international adoption. She reported no exposure to Korean after the adoption and no knowledge of it. All other participants had acquired English as their L2. Their age of acquisition (AoA) for English fell into two groups, one group with an AoA around age 6, the other group with an AoA around age 10. 17 participants had also learned German at some point in their life, usually as an L3 in high school and not earlier than age 13. 12 participants had learned additional foreign languages.

The L2 group consisted of 32 Norwegian L2 speakers with German as their L1. 17 participants were enrolled in the program for Scandinavian Studies at the Humboldt University Berlin. These 17 participants had also been instructed by the same teacher while learning Norwegian in Berlin. The remaining participants had acquired Norwegian in programs by other universities or during stays in

Norway, usually as part of an exchange program (high school, university or au pair). Four participants were residing in Norway at the time of testing.

These 32 participants (male n=14) had an average age of 26.19 years (range: 20-58, SD: 6.87). Two participants were left-handed. All had normal or corrected-to-normal vision and no history of language or neurological disorders. No participant had acquired an additional language before entering elementary school. For the majority of participants (n=17), Norwegian was the third foreign language (L4) and the fourth foreign language (L5) for another 9 participants. All participants had learned English before beginning to learn Norwegian and except for three participants, English was the L2. Average AoA for Norwegian was 20 (range: 14-29, SD: 3.1) and average length of learning Norwegian was 4.7 years (range: 0.5 – 12, SD: 3.1). 28 participants reported some time of residence in Norway, mostly during academic exchange programs. Average length of stay in Norway was 16 months (range: 0-84, SD: 20.0) with a median stay of 12 months corresponding to an academic year. These data indicate that for the L2 group Norwegian is a language that was acquired after puberty and mostly in a university setting. The motivation for learning Norwegian is the voluntary enrolment in the subject at university and/or the perspective of spending time in the country.

Due to the absence of a feasible placement or proficiency test, the participants in the L2 group were asked to rate their ability in the following tasks in Norwegian: reading, writing, speaking and listening. The scale was from 1 to 6 and the rating instructions told the participant that 1 corresponded roughly to level A1 of the Common European Framework of Reference for Languages (CEFR, http://www.coe.int/lang-CEFR) and 6 corresponded roughly to level C2. If participants were unfamiliar with the levels of the CEFR, the official level definitions were provided by the experimenter. The results for each area can be found in Table 4.3. The average score across all competences is 4.33 (or 17.31 out of a total of 24), which corresponds to the B2 level of the CEFR. The participants can be considered intermediate to advanced speakers of Norwegian by their self-rating. The group of students from the Scandinavian Studies program had to have reached B2 level in order to participate in the classes they were recruited from and were also judged by their teacher to actually have this level. The self-rating for the ability of reading that is essential for the experiments reported in this thesis was the highest of all abilities and the majority of participants judged their reading skills as better or equal to any other skill.

	M (SD)	range
Reading	4.97 (0.74)	3 - 6
Writing	4.03 (0.95)	2 - 6
Listening	4.34 (0.87)	3 - 6
Speaking	3.97 (1.03)	2 - 6
Overall - average - sum	 4.33/6 (0.72) 17.31/24 (2.89)	 3.25 – 6 13 – 24

Table 4.3 Self-rating data by task for the L2 Norwegian group, SDs in brackets

A full overview of the biographical data by participant can be found in Appendix C for both the native and the non-native group.

Materials

There were 12 critical sentences in this task. Six single sentences that had been discarded from the set of 24 final sentences for experiment 1b after the pretest were included in this task. They had received high plausibility ratings and were deemed fit for reuse in the agent identification task. An additional six SVO sentences were modeled after them using two animate NPs. All critical sentences were presented in the short version that had also been used in the pretest: NP1 – auxiliary – main verb – NP2 (for an SVO sentence). To reduce the repetition of the same verb *ville* (‘will, want’), three new auxiliaries were used: *måtte* ‘must’, *skulle* ‘shall’ and *kunne* ‘can’, both in present and past tense. This task had a 1x2 design as the factor word order was manipulated with two levels: SVO vs. OVS. NP animacy was not systematically varied, and subject and object appeared in three different animacy combinations. Half of the sentences showed no difference in animacy between object and subject as both were animate (81a,b). The other half showed a difference in animacy between subject and object, two sentences featured an inanimate subject with an animate object (82a,b), and four had an animate subject with an inanimate object. This opposition between two animate NPs on the one hand and one animate and one inanimate NP on the other is the one found in the experiments manipulating animacy that were reported in section 3.2.1.2. There was one within-items factor – word order that was expressed as either SVO or OVS order. NP animacy was a between-items factor with either equal animacy of the two NPs or unequal animacy. It did not lead to an increased number of experimental conditions.

SVO condition – equal animacy
(81a) Prinsen ville male prinsessen.
prince-the want to$_{PAST}$ paint princess-the
'The prince wanted to paint the princess.'

OVS condition – equal animacy
(81b) Prinsessen ville prinsen male.
princess-the want to$_{PAST}$ prince-the paint
'The princess, the prince wanted to paint.'

SVO condition – different animacy
(82a) Fiskegarnet vil omgi dykkeren.
fishing net-the will$_{PRES}$ surround diver-the
'The fishing net will surround the diver.'

OVS condition – different animacy
(82b) Dykkeren vil fiskegarnet omgi.
diver-the will$_{PRES}$ fishing net-the surround
'The diver, the fishing net will surround.'

The 16 filler sentences came from three categories: declarative main sentences with inversion, cleft sentences and sentences in present tense in which the syntactic roles could be assigned solely on the semantics of the verb. The four sentences with inversion were all XVSO sentences beginning with a temporal adverb followed by a verb-subject inversion. The eight cleft sentences had a topicalized NP followed by either a subject or an object relative clause. In subject clefts, the object is found after the main verb at the end of the sentence, while in object clefts the subject is found after the relative pronoun in front of the auxiliary and/or the main verb. Half of these fillers were object clefts. The four sentences with semantic disambiguation were present tense sentences without a modal and should therefore not be disambiguated through word order, but based on the semantics of the verb. Animacy of the subject and object was mixed in these sentences and half of them were OVS sentences. Altogether 16 of the 28 sentences in this task had a subject-first order, the remaining 12 sentences contained object-first order. This distribution was supposed to mimic the occurrence of OVS sentences in natural language, although a ratio of 4:3 is not the typical ratio, meaning that OVS sentences are fairly overrepresented in the experimental materials. The full set of all sentences used in this task can be found in Appendix A.

Procedure

One experimental session included the agent identification task reported in this section, the acceptability rating task from Experiment 3a (Section 7.3), and the SPR task including items from Experiment 1a and 3a. At the beginning of the experimental session, each participant filled out a questionnaire with biographical information and was handed an information sheet detailing the procedure of the following experiment, the handling of the data and the respect of personal information. Participants then gave their informed consent to participate in the experiment that could be withdrawn at any point. The L2 group additionally filled out a self-rating questionnaire assessing their language abilities in four areas of Norwegian languages use. This questionnaire was handed out before the actual experiment in order to avoid an influence of performance in the experiment on self-rated proficiency. The participants then completed the SPR task followed by a short vocabulary lists for the L2 group. Participants were asked to indicate any unfamiliar words. More detailed specifications for the procedure of the SPR task and the vocabulary list are given in Section 4.4. After the SPR task, participants completed the acceptability rating task reported in Section 7.3. The agent identification task was the last task of this experimental session. It was presented as an untimed pen-and-paper questionnaire. The instructions told participants they would be reading 28 sentences now and their task was to identify who is acting. As an example the simple SVO sentence *Kvinnen synger sangen* ('The woman sings the song.') was given, along with the question *Hvem synger?* ('Who sings?'). Then the correct answer was explained: *Da er det kvinnen som handler, fordi hun synger* ('It is the woman who acts here, because she sings') A simple SVO sentence was chosen as an example in order not to bias participants towards the auxiliary construction and OVS sentences. The term 'subject' was also omitted to avoid a potential bias for animate subjects or complications due to the use of psych verbs. The question was only used for illustrative purposes in the practice item to teach participants how to identify an agent and was not present in the experimental items. The sentences were written in Century Gothic font size 9 using a Word questionnaire template. The answers were given as binary choices with the order of the two NPs randomized to avoid a bias for a specific answer. To increase the salience of the two possible answers, they were printed in bold face. Figure 4.5 below shows the instructions, the example and first item of one list of the agent identification task. All sentences were grammatical and their order in the task was pseudorandomized. The critical sentences were spread across two lists using a Latin Square design. Two additional lists presented the items in a reversed order. The completion of this questionnaire took between 5 and 10 minutes. Participants were informed about the aim of the questionnaire after completion and were encouraged to provide feedback.

Du vil lese 28 setninger nå. Oppgaven din er å identifisere den som handler.
Eksempel:
Kvinnen synger sangen. -> Hvem synger?
Da er det kvinnen som handler, fordi hun synger.

1.___Jegeren skal løven drepe.

a) **jegeren** b) **løven**

Figure 4.5 Instructions, example and first item of the agent identifications task of Experiment 1a

The whole experimental session took between 30 and 45 minutes for the L1 group and between 45 and 60 minutes for the L2 group.

Predictions

The predictions for the main analysis center on the manipulation of the word order.

A – If the participants can successfully use the order of main verb and NP2 to assign thematic roles, then they should perform at ceiling in both the SVO and the OVS condition.

B – Øvrelid's (2004) corpus study had found about 10% of the sentences investigated were object-first sentences, making subject-first sentences the predominant word order. If the participants employ a subject-first strategy, they should perform at ceiling in the SVO condition, as the subject is in the first position in these sentences, and they should show low accuracy in the OVS condition as it does not fit with the subject-first order.

Strategy	Accuracy SVO	Accuracy OVS
A – main verb/NP2 order	++	++
B – subject-first	++	--

Table 4.4 Idealized accuracy patterns for Experiment 1a

Possible influences of the between-items manipulation of animacy will be explored in a post-hoc analysis.

C – Øvrelid's (2004) corpus study on non-auxiliary sentences had found that lower animacy of subjects and equal animacy of the two NPs block the topicalization of objects. If the same applies to the auxiliary sentences used in this task, a post-hoc error analysis could show higher error rates for the sentences with two animate NPs compared to those with only one animate NP. Sentences with inanimate subjects and animate objects might have a higher error rate than those with animate subjects and inanimate objects, which allow object topicalizations for non-auxiliary sentences.

L2 predictions

Generally, all of the above-mentioned predictions for the L1 group also apply to the L2 group. The following predictions are based on more L2-specific properties like proficiency and AoA that do not apply to L1 speakers. These factors have not been experimentally manipulated and participants had not intentionally been matched with regard to proficiency or AoA. The circumstances under which Norwegian is acquired (usually at university), however, lead to a very homogeneous group with regard to AoA and exposure time. Any noise that these measures could introduce into the dataset will be explored in post-hoc analyses. While the predictions below are based on results of previous research, the variation in the investigated L2 group might not be big enough to show clear effects.

D – A low, prepuberty AoA has often been found to be a good indicator of nativelike performance, while non-nativeness in performance increases with increasing AoA, especially beyond puberty (Abrahamsson, 2012). As my participants in this experiment had all started to acquire Norwegian after puberty, I expect little to no influence of AoA on the offline performance, even thought AoA ranged from 14 to 29 years (average 20) among L2 participants.

E – More proficient L2 speakers have been found to be more likely to achieve nativelike performance in offline tasks. Higher self-rated proficiency should then lead to a more accurate performance in the OVS condition. If the speaker has so far failed to acquire object topicalizations, he or she should show few correct answers in the OVS condition, and would likely show a strong subject-first preference. If non-target-like acquisition is limited to this one particular structure, performance on the non-canonical filler sentences should be unaffected.

F – Exposure-based accounts of L2 performance predict more nativelikeness with more exposure to the L2. I have two possible measures of exposure. One measure is the time spent studying Norwegian in years, (i.e. length of classroom exposure). Another measure is the time spent living in Norway in months (i.e. amount of exposure in an immersion setting). Exposure during immersion should be a better predictor of native-like performance, as it likely accounts for more exposure to the target structure than classroom exposure to the language in a non-immersive setting.

L2-specific property	**Effect on offline accuracy predicted?**
D – AOA	No
E – proficiency	Yes (higher proficiency -> higher accuracy)
F – exposure 1) non-immersive exposure 2) immersive exposure	 No Yes (longer immersion -> higher accuracy)

Table 4.5 Predictions for L2-specific measures on offline accuracy in Experiment 1a

Results

Both groups showed a similar general pattern in the agent identification task in that the OVS condition turned out to be more difficult than the SVO condition, resulting in much lower accuracy scores for the OVS condition. The average accuracy scores across both conditions and fillers for both groups can be found in Table 4.6 below. Overall accuracy was above 83% suggesting that the participants paid attention to the task. High accuracy scores for the SVO condition and the fillers also suggest that there was no fundamental attention problem for the two groups as a whole. As can be seen from the accuracy ranges below, some participants might have had attention problems; these participants will be dealt with in a post-hoc analysis.

Condition	L1	L2
SVO	95.81% (SD: 11.21, range: 50 - 100%)	99.47% (SD: 3.0, range: 83 - 100%)
OVS	65.13% (SD: 32.58, range: 17 - 100%)	65.13% (SD: 33.4, range: 0 - 100%)
Fillers	86.13% (SD: 7.73, range: 62.5 - 100%)	91.41% (SD: 8.9, range: 63 - 100%)
Overall accuracy	83.71% (SD: 9.51, range: 64 - 100%)	87.50% (SD: 10.1, range: 57 - 100%)

Table 4.6 Mean accuracy scores in the agent identification task (Experiment 1a) per group per condition, no participants or items excluded

A between-groups ANOVA run on the accuracy scores of the groups showed no main effect of Group ($F_1(1,62)=0.22$, $p=0.643$, $F_2(1,11)=0.4$, $p=0.54$) or interaction with Group ($F_1(1,62)=0.16$, $p=0.694$, $F_2(1,11)=0.63$, $p=0.44$), but a main effect of Order ($F_1(1,62)=61.23$, $p<0.001$, $F_2(1,11)=104.61$, $p<0.001$). This main effect is caused by the overall higher accuracy of SVO sentences than of OVS sentences. The OVS condition also shows a lot of variation in both groups, with a range from zero to 100% accuracy. This variance will be addressed in the following post-hoc analyses that will look in more detail at the different predictions.

Post-hoc analyses

As the performance of both groups in the OVS condition showed a lot of variation, I took a closer look at the individual participants and items to see if the variation could be explained by problematic items or unusual participants. The distribution of accuracy scores across participants is similar in both groups. 12 native and 11 L2 speakers reached 100% accuracy in the OVS condition, while 10 native and nine L2 speakers scored below 50%. A look at the overall accuracy showed that some of the participants with low accuracy in the OVS condition might not have paid close attention to the accurate solution of the task as they also showed low accuracy scores on the filler items. I re-calculated overall accuracy scores excluding participants who showed an overall accuracy of less than 70 across all sentences, or who reported not to know the OVS construction (two L1 speakers, two L2 speakers). However, this did not lead to a great improvement in accuracy in the OVS condition. The L1 group improved by 2% in both the SVO and OVS condition, up to 97.2% and 67.2% respectively, whereas the L2 group improved by 4.2% in the OVS condition up to 69.3% as they performed already at ceiling in the SVO condition.

A look at the individual items (Figure 4.6) revealed no item with an accuracy score of more than 2 SD above or below the mean accuracy for the L1 group, but one such item with an unusually low accuracy score for the L2 group. The item is 'The teacher, the plumber will help', and the comparatively low accuracy score could be due to a lexical effect: the L2 speakers might have considered the teacher to be the better prototypical helper than the plumber. Recalculating accuracy scores for the L2 group after removing this item (along with the previous exclusion of the abovementioned participants) leads to a further improvement of the OVS accuracy score up to 72.4%, an overall improvement of 7% from the original accuracy score. I did not exclude this item from the L1 group data, as it is the item with the fourth highest accuracy score in this group. However, removal of low-performing participants and unusual items did not lead to a change in the overall result. The accuracy in the OVS condition was still significantly worse than in the SVO condition, and the increase in accuracy in the L2 group did not lead to an interaction of condition with group.

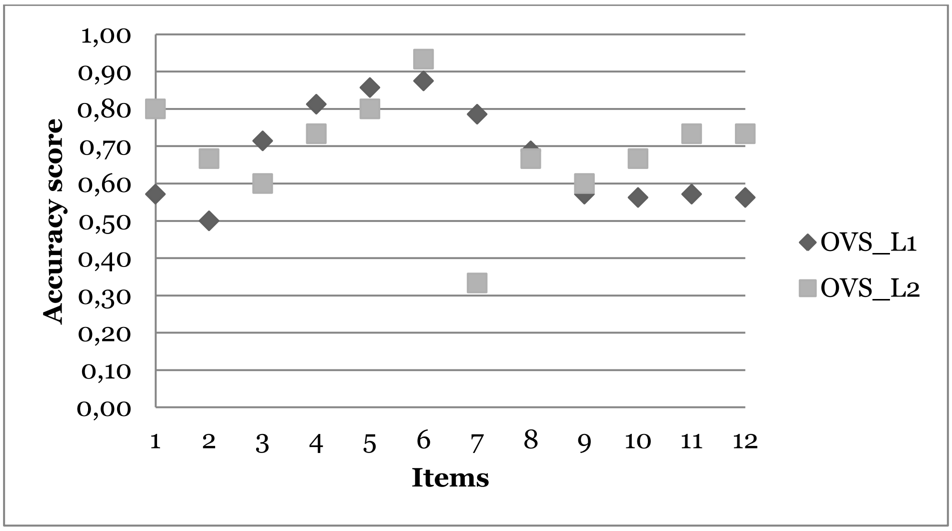

Figure 4.6 L1 and L2 accuracy scores by item in OVS condition only (Experiment 1a)

Apart from item 7 which shows a big difference in accuracy between the two groups, the performance of the two groups on individual items is similar. The L2 group performed better than or on par with the L1 group on items with inanimate subjects (1+2), and those items with two animate NPs (8-12). Sentences with an animate subject showed slightly higher accuracy scores for the L1 group (3-6).

	L1 group	L2 group
Inanimate subject Animate object (2 items)	71.88% (SD: 28.22, range: 0 - 100%)	84.38% (SD: 23.55, range: 50 - 100%)
Animate subject Inanimate object (4 items)	90.63% (SD: 13.84, range: 75 - 100%)	86.72% (SD: 16.78, range: 50 - 100%)
Two animate NPs (6 items)	76.56% (SD: 21.94, range: 33 - 100%)	78.65% (SD: 20.4, range: 50 - 100%)

Table 4.7 Accuracy scores by animacy setting for L1 and L2 groups (Experiment 1a)

Table 4.7 compares the accuracies for the different animacy settings across the two order conditions and shows a similar picture as Figure 4.6. The L1 group shows high accuracy scores for items with animate subjects and inanimate objects compared to a sharp drop in accuracy for inanimate subjects or sentences with two animate NPs. The L2 group shows a drop in accuracy when two animate NPs are involved, but when the animacy of the two NPs differs, their accuracy score is comparable.

I then carried out an error analysis. Table 4.8 gives the number of incorrect responses out of all responses and the percentage of errors for both groups separately and split according to sentence types.

Order	**Type of sentence**	**L1**	**L2**
SVO	Inanimate subject Animate object	1/30 3.33%	0/30 0%
	Animate subject Inanimate object	0/60 0%	0/60 0%
	Two animate NPs	4/90 4.44%	1/90 1.11%
OVS	Inanimate subject Animate object	14/30 46.67%	8/30 26.67%
	Animate subject Inanimate object	11/60 18.33%	14/60 23.33%
	Two animate NPs	34/90 37.78%	34[4]/90 37.78%

Table 4.8 Error analysis of L1 and L2 groups in agent identification task, amount of errors out of overall amount of answers (Experiment 1a)

An error in the first type of sentence (inanimate subject/animate object) meant choosing the animate NP instead of the correct inanimate one, in the second type

[44] 10 of these 34 errors are caused by item 7, the removal of this item leaves 24 errors in 75 sentences decreasing the error rate to 32%.

of sentence choosing the inanimate NP instead of the correct animate one, and in the third type of sentence choosing the wrong animate NP as the subject. The error analysis showed that the distribution of errors was different. The L1 group made more mistakes overall than the L2 group. In the OVS condition, sentences with two animate NPs were equally problematic for both groups. In fact, these were the sentences with the highest error rate in the L2 group. The L1 group had the highest error rate for sentences in which they incorrectly chose an animate NP over an animate one. When the two conditions with only one animate NP are summed up to be more comparable in number to the condition with two animate NPs (six items vs. six items), the conditions with only one animate NP show less errors (25/90, 27.78% in L1; 22/90, 24.4% in L2) than those with two animate NPs.

A look at the filler sentences revealed that the two groups showed similar trends in the inversion and the cleft sentences, but differed in their direction on the semantically disambiguated items. Fillers with inversion caused only occasional errors, 1/120 (0.83%) for the L1 group and 2/120 (1.67%) for the L2 group. A low error rate was also found for the cleft sentences with 3/120 (2.5%) for subject clefts and 9/120 (7.5%) for object clefts in the L1 group and 1/120 (0.83%) and 7/120 (5.83%) in the L2 group. These percentages were well below the ones for the OVS auxiliary sentences suggesting that participants did not have a fundamental problem with non-canonical word orders. Semantic disambiguation proved to be more difficult for the L1 group than the L2 group. The L1 group made 48 mistakes in 60 sentences, resulting in an error rate of 80%, while the L2 group only made 23 mistakes (38.3%). A mistake in these OVS sentences was either choosing an animal over a human as the agent or a human over an inanimate entity as the agent of the psych-verb. The accuracy rates for each experimental and filler item in the agent identification task are reported in Appendix B.

In order to investigate the influence of AoA, proficiency and exposure on the accuracy of the L2 group in the OVS condition, I ran separate linear regressions. I did not include the SVO condition as all but one participant had performed perfectly in this condition. There was no effect of AoA on the accuracy scores, $p=0.55$ and participants across the entire range of AoAs were able to perform at ceiling.

The linear regression with proficiency as a predictor variable used each participant's summed up proficiency score across all four self-rated abilities, giving a larger range of proficiency scores than just using the rating for the reading ability. Still, there was no significant effect of proficiency on the OVS accuracy scores, p=0.19. Figure 4.7 below shows the scatter plot for overall proficiency and OVS accuracy scores. Two data points seem to be outliers with high self-rated proficiency in the face of very low accuracy scores, suggesting a possible overestimation of language abilities. Removing these two data points from the analysis results in a significant influence of proficiency on the accuracy scores, p=0.013.

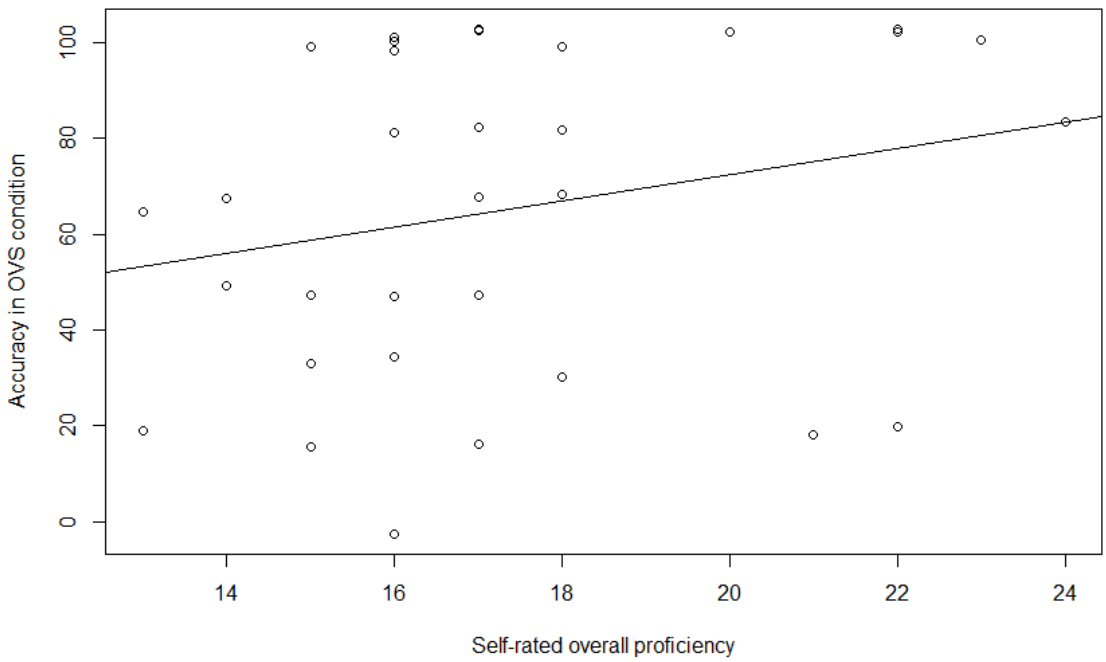

Figure 4.7 Self-rated overall proficiency and accuracy scores in the OVS condition (Experiment 1a)

The two measures of language exposure yielded better results, as the time spent learning Norwegian approached significance (p=0.069) and the time spent living in Norway was significant (p=0.02). Figure 4.8 below compares the scatter plots for time learning Norwegian and time living in Norway.

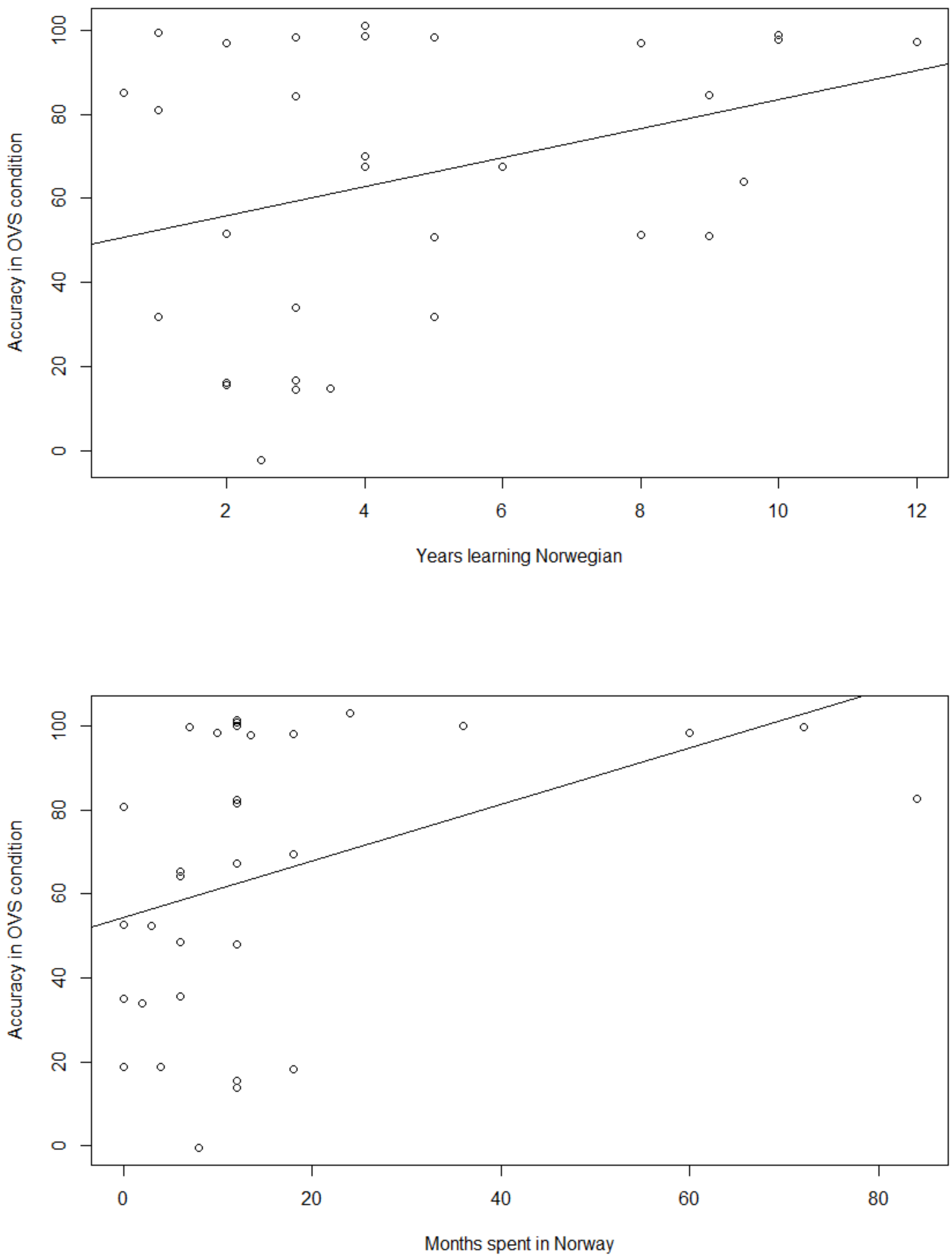

Figure 4.8 Comparison of time spent learning Norwegian and time spent living in Norway by accuracy scores in the OVS condition

Both plots show that longer exposure was associated with a better performance, although learning Norwegian for a prolonged period of time, e.g. eight years, did not guarantee above-chance accuracy and participants with much less exposure also performed at ceiling. The same analyses for overall accuracy only showed an effect for time spent in Norway, suggesting a generally beneficial effect of immersion on learning the target structure.

Summarizing the results, it can be said that non-canonical word orders were not difficult per se as both participant groups dealt well with sentences with an inversion and object cleft sentences. Even though both groups had highly similar accuracy scores in the task, an error analysis showed that they differed in the relative frequency of different error types. The L1 group made more mistakes based on NP animacy than the L2 group.

Discussion

The agent identification task showed that OVS sentences with auxiliaries are hard to comprehend for native and non-native speakers alike, even when given enough time to consider their response and possibly even reflect on syntactic rules. The data from filler sentences containing other non-canonical word orders suggest that there was no fundamental problem with non-canonical word orders per se as performance on the fillers was at ceiling. The final overall accuracy scores of the L1 and L2 group were nearly identical and did not correspond to either prediction A (success for OVS structures) or B (failure for OVS structures). Instead, there was a lot of variance in both groups and the full range of accuracy scores was present. There also seemed to be an influence of animacy on the assignment of agency in the L1 group that was less prominent in the L2 group as evident by the different errors made by the two groups. The main questions to be discussed now are the following: What makes object topicalizations in auxiliary sentences different from other non-canonical word orders and why are they more difficult to comprehend in an offline task? How can the wide range of accuracy scores in both experimental groups be explained and what are possible motivations for the difference in errors between the L1 and the L2 group?

Object topicalizations in auxiliary sentences are different from the other non-canonical structures investigated as fillers in that they are less frequent and the cue that signals object topicalization is unique to these structures and very subtle. There are no additional syntactic structures highlighting a possible change to the information structure as in cleft sentences with their introductory *Det er…* 'It is…' phrase. Object cleft sentences also use the same changes to the order of

NP and main verb that are used in object relative clauses, essentially because they are object relative clauses after the cleft. The frequency of occurrence for this type of object topicalization is higher than for auxiliary sentences. In a task like this written agent identification task, participants assign stress to the sentence while reading and topicalized objects usually require contrastive stress. While the cleft structure unambiguously signals stress on the clefted element from the start of the sentence, the auxiliary sentence signals this change to sentence prosody only later in the sentence, requiring a more effortful revision of sentence prosody. As had been demonstrated by Kristensen, Engberg-Pedersen, & Poulsen (2014) in their study on Danish object-initial sentences, an appropriate context improves comprehension of object topicalizations in auxiliary sentences (from 51 to 75%); furthermore, the actual presence of object topicalizations are explained by their context. In actual discourse, sentences rarely appear out of context, which helps the comprehension of object topicalizations. In this agent identification task, object topicalizations appeared without any context. While the cleft construction strongly signals a topicalization, there was no motivation to assume a topicalization for the auxiliary sentences. Frequency and additional syntactic structure that immediately signals a topicalization made cleft structures a lot easier to comprehend and helped in agency assignment compared to auxiliary sentences.

The order of main verb and NP2 in an auxiliary sentence is apparently not as reliable for sentence comprehension and agency assignment as has been claimed by the Norwegian Reference Grammar (Faarlund et al., 1997). Taking a CM perspective offers a possible explanation as to why this is the case. As already mentioned the main verb/NP2 order cue is not very frequent and only applies to this very specific auxiliary structure resulting in low cue availability. Additionally, object-first structures are comparatively infrequent in Norwegian. Øvrelid (2004) found a 10% occurrence in her corpus of ambiguous non-auxiliary sentences and the frequency of object-first auxiliary sentences in the entire Norwegian language will be even lower. This means that the task of assigning an OVS structure to any kind of sentence has a low task frequency resulting in low cue strength. The cue itself is a global cue involving two adjacent elements and the positing of a filler-gap structure thereby affecting the sentence structure in its entirety. This is associated with high demands on working memory and low cue assignability. The change in order of main verb and NP2 is also possibly not very salient - a cleft structure would be a more salient alternative - meaning the cue also suffers from low perceivability. Low assignability and low perceivability mean that the cue comes with high costs that decrease its reliability and validity. The statement of the Norwegian Reference Grammar clearly assumes a high conflict reliability of the main verb/NP2 order cue that is not reflected in my results. Conflict reliability of a cue is the hallmark feature of adult sentence processing in the CM framework. While children rely more on cue availability, adults use conflict reliability to estimate cue strength: the more reliable a cue is in conflict, the stronger it is (MacWhinney, 2005). In theory, the main verb/NP2 order cue should be maximally reliable in conflict and trump all other cues like NP animacy or the sentence-initial position and it does, but only for a subsection of all participants. Assuming that the L1 participants went through the same L1 acquisition process as children and were exposed to highly similar frequencies of the individual cues, where does this difference come from? The high cue cost associated with its demands on working memory and its possibly low saliency is what affects the cue's conflict reliability and its strength. The only way to decrease its cost and strengthen its position in the cue hierarchy is to be more salient or less demanding for working memory. Individual differences in working memory capacity between the participants could be an explanation at this point. Previous research has found differences in ERPs patterns between speakers with high and low working memory during the processing of filler-gap dependencies and suggesting less difficulty to keep the filler in active memory for speakers with high working memory scores (Fiebach, Schlesewsky, & Friederici, 2001). Native speakers with lower working memory scores have also been found to show slower reading times, lower comprehension accuracy scores and to benefit more from supporting contexts than speakers with high working memory scores (J. W. King & Just, 1991). Additional evidence also suggests an influence of working memory on the ability to maintain several parallel analyses at one time (MacDonald, Just, & Carpenter, 1992). Processing of the main verb/NP2 cue could be less demanding for participants with a higher working memory capacity, thereby increasing its assignability, reducing cue cost and boosting cue strength. I do not assume that participants with lower working memory capacity are completely unable to use this cue, but they might be less consistent in its application and depend more on the actual sentence in question and the influence of other cues. As working memory of the participants was not assessed in this experiment, this is highly speculative and only a follow-up study that involves the assessment of working memory and/or other cognitive control measures could provide evidence for or against this hypothesis. Another possibility to boost cue strength for the main verb/NP2 order cue could be the introduction of context, as has already been found to be beneficial in other studies. Providing a contrastive context should make participants more aware of the possibility of an object topicalization and possibly activate the use of cues that are specific to this construction (like the main verb/NP2 order cue), even in speakers in which these cues are normally given less priority in agency decisions.

The main verb/NP2 order cue did not occur in isolation and other cues clearly played a role in the agency assignment of a majority of speakers that completed the task with high error scores in the OVS condition. NP animacy and the sentence-initial position were two other cues that speakers could and did make use of. I will discuss animacy first. The influence of NP animacy in the L1 group is not surprising based on the corpus study by Øvrelid (2004) and the results in other languages like Russian that had found an influence of animacy despite unambiguous morphological case marking (Stoops et al., 2014). Øvrelid had found two contexts in which animacy blocks object topicalization in ambiguous NP1 V NP2 sentences. There was no object topicalization when the object was higher in animacy than the object or when they were equally animate. The agent identification task featured three combinations of animacy: subject higher in animacy than object, subject lower in animacy than object, subject and object both animate. The latter two combinations would block object topicalization in ambiguous sentences. This rule seems to affect agency assignment also in unambiguous auxiliary sentences. Taking the overall accuracy for each animacy combination, the one combination that allows object topicalization also in ambiguous structures (animate subject/inanimate object) reached an overall accuracy of 91%, while the two remaining combinations that would be blocked in ambiguous structures reached much lower accuracy scores: 72% (inanimate subject/animate object) and 77% (both animate). The same trend was visible when only focusing on the errors in the OVS condition. Blocked combinations had error scores as high as 47% (inanimate subject/animate object) and 38% (both animate), while the animate subject/inanimate object combination was less problematic, with only 18% errors. Although the main verb/NP2 order cue enables object topicalization for any combination of animacy, it seems that a sizeable portion of L1 participants ignored this cue. Instead, they applied the same animacy contexts for object topicalization to unambiguous sentences that would allow object topicalization also for ambiguous sentences, i.e. for these speakers, object topicalization would only be possible when the subject is animate and the object inanimate independent of the syntactic context.

It could be claimed that the L1 participants simply followed an agent-first strategy, when they did not successfully apply the main verb/NP2 order cue. Based on the difference in accuracy scores for the different animacy combinations, this strategy would not be a pure agent-first strategy, but modulated by NP animacy. An agent-first strategy would have lead to no errors in the SVO condition; there were, however, occasional mistakes in the SVO condition. These rare errors occurred in sentences with two animate NPs and with an inanimate subject – the same sentence types that had shown elevated error rates in the OVS condition. While a tendency to prefer subjects in the first position is undeniable, it might be subordinate to the strong association of agency with animacy. Sentences involving two animate NPs always had an animate NP in the sentence-initial position in the OVS condition, as did OVS sentences with an animate object and an inanimate subject. In both cases choosing the first NP as the agent would encourage the association of animacy with agency. Both sentence types would also block object topicalization in ambiguous sentences. Nevertheless, the L1 participants made more errors in sentences with inanimate subjects and animate objects than in sentences with two animate NPs. Clearly, having two animate NPs introduces a competition for agenthood that is not present with a second inanimate NP, although the second animate NP is not in a sentence-initial position.

Based on the error rates for the different animacy settings and the two order conditions, I suggest the following hierarchy in (83) for the three cues NP animacy, main verb/NP2 order and subject-initial position for the Norwegian L1 group as a whole[5]:

(83) animacy > main verb/NP2 order > sentence-initial NP

Even though NP animacy was only a between-items factor, I suggest ranking it as the dominant cue as it has an influence on all other cues, and the results of the agent identification task cannot be explained without recursion to NP animacy. Animate subjects that were the only animate NP in a sentence had higher accuracy rates independent of the order manipulation. Animate objects paired with inanimate subjects caused high error rates in the OVS condition, suggesting that the initial decision to assign agency to the sentence-initial animate NP was not revised. When the animacy cue was neutralized in the sentences with two animate NPs, it was not the sentence-initial NP that attracted the agency decisions, instead there were more correct answers based on the order of main verb and NP2. This allows the cue to outrank the sentence-initial position cue. For the subset of participants that performed flawlessly on this task, the ranking needs to be different with main verb/NP2 order as the strongest cue that even outranks animacy, as their answers were not affected by animacy and likely solely driven by the main verb/NP2 order cue. A strict application of this cue always leads to an unambiguous answer and recursion to a secondary cue is not necessary.

This uncertainty with regard to the cue hierarchy in Norwegian stresses the need for more elaborate follow-up studies using an agent identification task to gather more evidence for one or against the proposed hierarchy. It also highlights a weakness in the CM framework. In the standard agent identification task that is normally used, the main verb/NP2 order cue would never be present as only NP1 V NP2 sentences are used, similar to Øvrelid's corpus study. Any extension beyond sentences of this type involve cues that are not present all the time and might still overrule all other cues like the order cue used in this study, and it is unclear how to accommodate these cue in the hierarchy. Is the cue simply muted when not present in the input or do speakers have different hierarchies at their

[5] NP definiteness was also found to play a role in Øvrelid's corpus study, but as it was not systematically manipulated in this experiment, I will not include it in my hierarchy. This means that the hierarchy suggested by me is only tentative and requires more testing to accommodate other cues.

disposal depending on the cues present in the sentence? The CM puts a lot of emphasis on the actual mathematical computation of cue strengths, and the presence or absence of a cue likely has an influence on the strength of the other cues. If a cue is muted or deactivated because it is not present in the current input what happens to its strength? Does only the next cue in the hierarchy increase in strength, or do all the cues' strength values change? What happens if the cue is situated in the middle of the hierarchy like in my suggested hierarchy in (83)? If we assume two different hierarchies depending on syntactic context (a possibility that is never discussed in the CM framework due to its limitation to simple structures), do speakers have two (or more) values of cue strength, reliability etc. associated with each cue that are activated depending on context? This would require an adjustment of cue values that takes place at the same time as a possible syntactic reanalysis, when the speaker realizes that a different hierarchy is needed. This would put an additional burden on the processing system and would not be very economical.

When describing their motivation to conduct an experiment to investigate the influence of animacy on the processing of non-canonical SOV+V structures in Russian, Stoops et al. (2014) suggest that '*different word orders in a given language might induce different parsing strategies, or at least serve to direct more attention to one factor and away from another.*' (Stoops et al., 2014:586) This attentional or strategic shift could be motivated by an alleviation of demands on working memory within a memory-based parsing framework. Similarly to the notion of conflict reliability in the CM framework, they suggest that unambiguous and thus more reliable information sources are weighted more strongly than less reliable ones. In their view, animacy information would be a more reliable and more easily retrievable cue than morphosyntactic information, which might not be well differentiated and might be difficult to reactivate. While limitations on working memory might play a weaker role in the untimed agent identification task reported in this chapter compared to a self-paced reading task as reported in Stoops et al., differences in memory resources might play a role in the detection of the cues as already discussed above. Only some participants might have recognized that a shift towards a different source of information, in this case away from animacy and towards word order, is needed. The rest of the participants possibly overemployed the 'regular' setting with a strong focus on animacy that is successful for ambiguous NP1 V NP2 sentences which block object topicalizations whenever the animacy criterion is not met. These participants did not pick up on the cue provided by the order of main verb and NP2 which signaled the need to adjust their focus leading to mistakes in agency assignment. I would not go as far as Stoops et al. and assume two entirely different processing routines for SVO and OVS sentences within the same speaker. I would rather assume the possible existence of different cue hierarchies within the same native speaker population that vary according to individual differences between the speakers. This difference has not been found previously as the structures investigated were not taxing enough to native speaker processing systems, and were restricted to globally ambiguous sentences in which there simply is not an incorrect answer.

Another theoretically possible explanation for the weak performance of some participants could be the 'good enough' approach by Ferreira and colleagues (2002). Participants would not have computed a fully specified syntactic representation of the OVS sentences, but instead computed a rough representation containing the action and the two NPs involved, but no strong commitment regarding agency during the reading of the sentence. The agency assignment would then be based on simple strategies such as a subject-first strategy, or an association of animacy and agency. There was no punishment or direct feedback on agency assignment, and an incorrect 'good enough' representation did not receive any other conflicting evidence (except for the order of main verb and NP2) that would force a reanalysis. As this was one of the last tasks after the self-paced reading task, some participants might have resorted to 'good enough' processing due to deficits in attention and exhaustion. Counterevidence for this point of view comes from the successful assignment of agency in cleft structures that would also have shown elevated errors under 'good enough' processing. It is unlikely that participants resort to shallow or 'good enough' processing only for particular structures.

Most of what has been discussed for the native speaker group could also be applied to the results of the L2 group, but there were also marked differences between the L1 and L2 group that need to be discussed. The L2 group performed on par with the L1 group when looking at the overall accuracy scores, but they showed less sensitivity to the differences in animacy between the NPs. Two animate NPs as possible agents were the most difficult combination for the L2 group, but there was no difference in accuracy between the conditions with only one animate NP, suggesting that for German L2 speakers of Norwegian an inanimate NP is as good an agent as an animate NP. There was also no indication of a subject-first strategy. The L2 group seemed to have been guided more by the semantics of the individual verbs which would also explain their much better performance on the semantically disambiguated filler sentences compared to the L1 group. Errors in the L2 group were spread out more across the different items than in the L1 group, with the exception of one item that emerged from the post-hoc analysis as having an unusually high error rate.

The CM framework assumes that when speakers acquire a foreign language, they start out with the cue hierarchy and settings of their native language and gradually adjust them to fit the native pattern of the L2 (MacWhinney, 2005). The German cue hierarchy suggested by MacWhinney can be seen in (84):

(84) case > animacy > agreement > word order

As case and agreement are not available, the initial state of the L2 speakers' cue hierarchy in Norwegian would simply be animacy > word order. The absence of an overwhelming effect of animacy is evidence that the L2 group investigated had already moved beyond this initial state and had apparently begun to integrate the additional main verb/NP2 order cue into their system. With the exception of two participants that reported not to know this cue and also had problems with other non-canonical constructions, all other participants were able to at least partially use the main verb/NP2 order cue in their assignment of agency. A possible explanation for the variation within the L2 group could be that the individual participants rely differently on the information provided by word order, animacy and the individual verbs and make their decisions based less on a general hierarchy, but more on a case-by-case basis. Animacy seemed to be more influential or distracting when there were two animate entities than when there was just one.

The analysis of L2-specific factors on accuracy yielded no surprising results. The absence of effects of AoA on the accuracy scores can be explained by the fact that the L2 group was comparatively homogenous with regard to the AoA, as all participants had started to learn Norwegian after puberty, and strong effects of AoA are mainly found when comparing groups with AoAs before and after puberty. The range of AoAs from 14 to 29 years might seem big at first and a possible source of variance, but the vast majority of participants clustered around the AoA of 19 marking the beginning of university studies. It has to be kept in mind, however, that this group with comparatively late L2 learners performed like the L1 group when only absolute scores were considered. Roughly the same number of L1 and L2 participants performed below chance and at ceiling. The analysis of the two measures of exposure suggest that prolonged immersion as measured by the months spent living in Norway was the better indicator of higher accuracy of these two measures, as all of the participants with an immersion of several years reached high accuracy scores. But the number of those participants was comparatively small, and some participants with significantly shorter time spent in Norway were also able to achieve good accuracy scores. More participants with prolonged immersion through residence in Norway would be needed to confirm this trend as the majority of participants had resided in Norway between 6 and 12 months. Time spent learning Norwegian in a classroom setting was less of an indicator for accuracy in the OVS condition. These results support exposure-based accounts of L2 processing that assume more native-like performance with prolonged exposure, but even though longer exposure always lead to higher accuracy scores, it was not impossible to reach a perfect score with comparatively little exposure. The results of the analysis for proficiency with a non-significant result for the full L2 group, and a significant result with two participants with spuriously high self-ratings removed, hint at the slightly problematic nature of self-ratings. While AoA and exposure measures are objective, self-ratings are subjective and although comparisons in the literature have found a high correlation between self-ratings and more objective proficiency measures, over- and underestimations of language abilities can and do occur. The effect found after the removal of two participants is in line with expected results of a better performance with higher proficiency.

Summing up, object topicalization in auxiliary sentences was a challenging structure for native and non-native speakers alike and only a subgroup of speakers was able to reliably assign agency in object-first conditions. When looking at the broad, overall picture based on average accuracy scores, there was no difference between the L1 and the L2 group. Differences only appeared after a closer inspection of error rates, revealing a differential effect of NP animacy in the two groups. Native speakers had a harder time choosing the correct agent when NP animacy did not fit their preferred setting (i.e. only one animate NP that is judged as the agent). The L2 group only showed elevated error rates when two animate NPs were present. In the following self-paced reading task, the animacy of the two NPs always differed. If the processing used to parse the sentences is the same in both tasks, native speakers should show strong effects in the OVS condition with an animate object and inanimate subject, which also caused high error rates in the offline task, and weaker effects in the OVS condition with an inanimate object and animate subject. If the online processing of L2 speakers is similar to their offline performance with regard to the influence of animacy, there should not be a difference between the OVS conditions, as there was also no difference in accuracy between the two animacy settings in the offline task.

4.4 Experiment 1b: Self-paced reading task

The self-paced reading task investigated the online integration of the two cues animacy and the NP2/main verb order that were found to be used differently by participants in the agent identification task. While participants had plenty of time to process the cues in the previously reported agent identification task, the self-paced reading task put more time pressure on their processing system and monitored real time word-by-word processing of the participants.

Participants

The participants were the same as described under Section 4.3.

Materials

The self-paced reading task employed a 2x2 design with word order (SVO vs. OVS) and subject animacy (animate vs. inanimate) as factors resulting in four experimental conditions. The 24 experimental sentence pairs that had been chosen based on their good plausibility result in the pilot study were modified for their use in the self-paced reading experiment (see Appendix A for a full list). One modification was the introduction of OVS sentences by swapping the position of the main verb and the NP2. The second modification concerned the insertion of additional lexical material. After the NP1, a relative clause was inserted to prolong the time until the disambiguation point and make participants stick longer to their interpretation of the first NP as the subject of the sentence. This relative clause consisted of the relative pronoun *som* 'who/that', an adverb, the verb *er* 'is/are', and an adjective. The exact order of adverb and verb depended on the type of adverb. The order for a sentential adverb can be seen in examples (85a-d) below. The adjective used was usually kept constant across all four conditions. In ten cases this was not possible due to the change in animacy of the NP. After the relative clause the sentence continued as in the pilot study with the auxiliary *vil* 'will', the main verb and the second NP. The order of the main verb and the second NP depended on the word order manipulation. Four main verbs were used twice in the experimental set of 24 sentences: *passere* 'pass', *berøre* 'touch', *treffe* 'hit' and *forandre* 'change'. In order to move the point of manipulation away from the end of the sentence and therefore control for possible spillover and sentence wrap-up effects, an adverbial of time or location was attached to the end of the sentences. In the following examples, the subject is printed in bold. Instead of sentence pairs used in the pilot study, one experimental item no consisted of a sentence quadruplet.

SVO order – animate subject (SVO_animate)
(85a) **Soldaten** som vanligvis er korrekt, vil lokalisere radaren i mørket.
Soldier-the who usually is$_{PRES}$ correct, will$_{PRES}$ localize radar-the in darkness-the
'The soldier who is usually correct will localize the radar in the dark.'

SVO order – inanimate subject (SVO_inanimate)
(85b) **Radaren** som vanligvis er korrekt, vil lokalisere soldaten i mørket.
Radar-the that usually is$_{PRES}$ correct, will$_{PRES}$ localize soldier-the in darkness-the
'The radar that is usually correct, will localize the soldier in the darkness.'

OVS order – animate subject (OVS_animate)
(85c) Radaren som vanligvis er korrekt, vil **soldaten** lokalisere i mørket.
Radar-the that usually is$_{PRES}$ correct, will$_{PRES}$ soldier-the localize in darkness-the
'The soldier will/wants to localize the radar that is usually correct in the darkness.'

OVS order – inanimate subject (OVS_inanimate)
(85d) Soldaten som vanligvis er korrekt, vil **radaren** lokalisere i mørket.
Soldier-the who usually is$_{PRES}$ correct, will$_{PRES}$ radar-the localize in darkness-the
'The radar will localize the soldier who is usually correct, in the dark.'

In order to check whether the L2 group knew the critical nouns and verbs, a vocabulary test was created. It featured low-frequency nouns, adjectives and verbs that had been used in the experimental sentences. The low-frequency cut-off for nouns and adjectives was set to less than 100 occurrences in the OCTT and to less than 500 occurrences for verbs. Several of these low-frequency words were nearly identical in orthography to their German counterparts. To reduce the number of items on the vocabulary test, three German native speakers without any knowledge of Norwegian were asked to check the vocabulary test for words they could identify. If all three of them could identify a word, it was taken off the list. The final vocabulary list included 61 words, 34 words came from Study 1 and 27 words came from Study 3.

During the self-paced reading task, participants were also given eight comprehension questions probing their final interpretation of the sentences. Half of the eight questions expected a negative answer, and half of the eight questions were formulated as passives. The questions were asked after those items that were based on sentence pairs with the closest plausibility rating in the pilot study. As the interpretation with an animate and an inanimate subject were nearly equally plausible, answers based purely on plausibility were deemed less likely than for sentences with bigger difference in plausibility.

Each experimental list began with four unrelated practice items and one comprehension question. There were 48 filler sentences, 24 of which were items for the Norwegian experiment on particle verbs (Section 7.3). The other 24 sentences were true fillers. Eight were pseudo fillers modeled after items of this experiment with either only animate or only inanimate NPs (86 and 87), another eight contained grammatical subject or object shift and were pseudo fillers from Study 3 (88 and 89). The remaining eight fillers represented various morphological and syntactic phenomena of Norwegian (see example 90). Eight questions also tested the comprehension of these fillers. Altogether, the participants saw 76 sentences and 25 comprehension questions.

(86) Tyven som er utrolig raffinert, vil bestjele røveren uten problemer.
thief-the who is$_{\mathrm{PRES}}$ incredibly elaborate will$_{\mathrm{PRES}}$ steal robber-the without problems
'The thief who is incredibly elaborate will steal from the robber without problems.'

(87) Bussen som er veldig koselig, vil krysse elven med fergen.
bus-the that is$_{\mathrm{PRES}}$ very cosy will$_{\mathrm{PRES}}$ cross river-the with ferry-the
'The bus that is very cosy will cross the river by ferry.'

(88) Kristoffer drikker gjerne øl, men Ida drikker det ikke.
Kristoffer drink$_{PRES}$ voluntarily beer but Ida drink$_{PRES}$ it not
'Kristoffer likes to drink beer, but Ida does not drink it.'

(89) Om morgenen vanner Liv aldri blomstene, men hun gjorde det i morges.
in morning-the water$_{PRES}$ Live never flowers-the but she do$_{PRET}$ it in morning
'In the mornings, Liv never waters the flowers, but she did it this morning.'

(90) Hun synes at hybelen hennes er ganske dyr.
she thin$_{PRES}$ that room-the her is$_{PRES}$ very expensive
'She thinks that her room is pretty expensive.'

The items were organized in three blocks of 24 sentences and pseudorandomized using the software provided by www.random.org/lists with an equal number of comprehension questions in each block. It was checked that no condition was overrepresented in any of the blocks. Items were assigned to four presentation lists based on a Latin Square design. Reversing the lists with the software provided by www.textmechanic.com/sort-text-lines.html resulted in altogether eight presentation lists. There were also special lists for left-handed participants that included a reversal of the answer buttons for the comprehension questions.

Procedure

The SPR task was the central part of the experimental session outlined in the procedure section of Section 4.3. It was run on an ACER notebook with a screen refresh rate of 16.72 ms and a 15.4" screen. DMDX software (Forster & Forster, 2003) was used to present the experimental items and record the reading times and responses to the comprehension questions. An Xbox 360 controller was used as input device. Altogether four different buttons were used in this experiment. The experimental sentences all began in the middle of the screen on the left side, starting with an >. Upon button press, the entire sentence was first presented as underscores. By pressing the green A button, participants could request the individual words of the sentence in a moving-window fashion. After a word was read, it was again shown as underscores. For clarification purposes, Figure 2.3 from the methods section is repeated here.

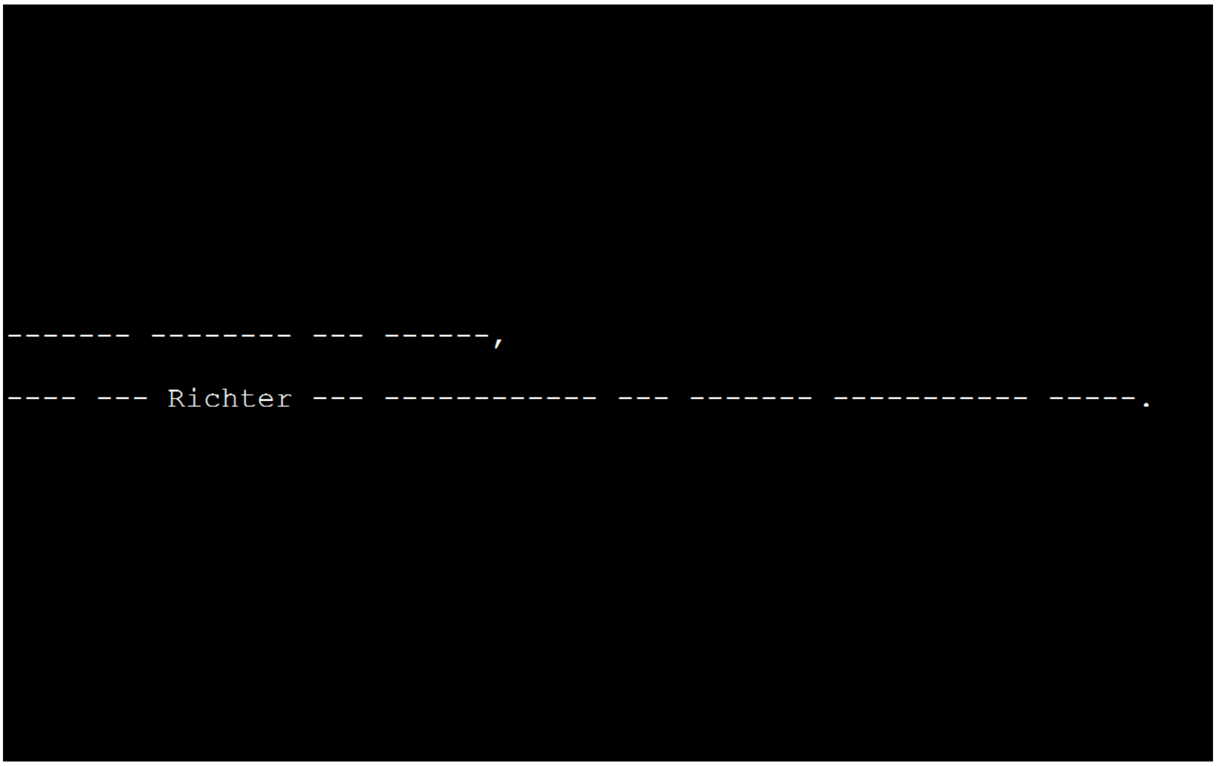

Figure 2.3: Screenshot from the SPR task in Experiment 2b

Some sentences had to be shown in two lines with the second line beginning directly below the first on the left side of the screen. The line break never occurred within the area of manipulation. The next sentence was requested by pressing the red B button. Answers to the comprehension questions were given by pressing either the left or the right bumper button. The button for a positive answer was always the same as the dominant hand of the participant. No feedback was given regarding the accuracy of the responses and the response time was measured from when the question was displayed, thereby including the time it took the participant to read the question. There was a timeout after 40 seconds, when the next trial was presented. Instructions told the participants to be as quick and accurate as possible in their reading and answering. They had the possibility to take two breaks during the experiment, after one and two thirds of the trials respectively. The experimental items and instructions were shown in Courier New font, size 11, white color on a black background. Encouraging comments during the experiment were shown in green. Comprehension questions were shown in Tahoma font, size 12 and red color. Reading time data was analyzed using repeated measures between-groups ANOVAs. The self-paced reading task took 10-20 minutes for the L1 group and 20-30 minutes for the L2 group.

Predictions

The predictions for the self-paced reading task are based on the two cues given in the task - main verb/NP2 order and NP animacy – and the possible application of these cues independently or in combination. The critical region that was analyzed contained the main verb and the NP2. Idealized reading times based on the predictions can be found in Figure 4.9 below.

A – If the participants can successfully use the order of main verb and NP2 to identify the non-canonical OVS order, then the participants have to abandon their initial interpretation of the NP1 as subject and reinterpret it as an object in OVS sentences. This syntactic reanalysis should result in generally longer reading times in the critical region for OVS sentences as compared to SVO sentences.

B – If participants apply the Animated First Principle as suggested by Tomlin (1986), they will prefer the order animate NP1 > inanimate NP2 independent of the thematic roles of the NPs. The two conditions with the order inanimate NP1 > animate NP2 will violate this principle and cause slower reading times.

C – If animacy and thematic role assignment are connected as suggested by Primus (1998), participants use both cues NP animacy and main verb/NP2 order to assign agency and solve the ambiguity. I then expect the fastest reading times for SVO sentences with animate subjects, as both cues have the preferred value. OVS sentences with inanimate subjects should have the slowest reading times, as both cues have the dispreferred value. The remaining two conditions each have one cue with a preferred and a second cue with a dispreferred value, resulting in reaction times in between the two extremes.

D – As SVO is the more frequent and canonical word order, processing might proceed more effortlessly in this condition as opposed to the OVS condition. Any effects of NP animacy might be masked in the SVO condition due to the absence of a syntactic reanalysis and overall high reading speed. In the OVS condition on the other hand, the additional information provided by NP animacy could lead to a quicker reanalysis for OVS sentences with animate subjects as opposed to OVS sentences with inanimate subjects.

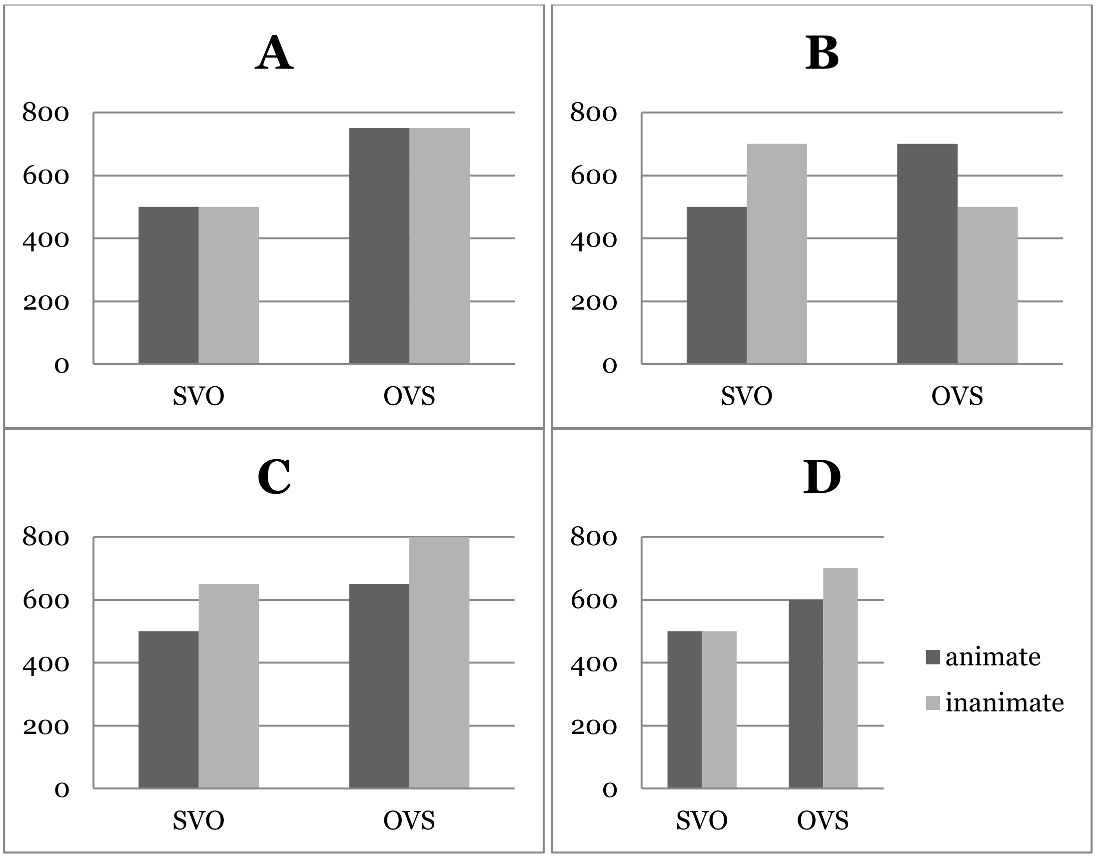

Figure 4.9 Idealized reading time patterns for the SPR task (Experiment 1b)

Results

Comprehension questions

Before analyzing the SPR data, I looked at the comprehension questions to see if any participant showed unusually low accuracy that could be related to decreased attention during the experiment. Accuracy scores for each individual question (experimental and filler) can be found in the corresponding materials section of Appendix A. In Experiment 1b participants had to answer eight comprehension questions, one after 1/3 of all experimental items. Another eight comprehension questions followed items from Experiment 3b and will be discussed in the results section in Section 7.4. A further eight comprehension questions followed filler items. Table 4.8 below shows the overall accuracy including the questions from Experiment 3b and the data from all participants.

	L1 group (N=32)	L2 group (N=32)	Both groups
Overall	70.05% (0.1) (range: 50-95.8%)	72.53% (0.1) (range: 54.2-91.7%)	71.3% (0.45)
Experiment 1b	63.67% (0.23) (range: 25-100%)	66.02% (0.21) (range: 25-100%)	64.84% (0.22)
Experiment 3b	94.14% (0.08) (range: 75-100%)	91.8% (0.09) (range: 75-100%)	92.97% (0.09)
Fillers	52.34% (0.13) (range: 25-87.5%)	59.77% (0.16) (range: 25-87.5%)	56.05% (0.15)

Table 4.9 Mean accuracy rates to comprehension questions Experiment 1b and 3b, SDs in brackets

As can be seen from Table 4.9, overall accuracy varied a lot between participants, from 50% to 95.8% in the L1 group and from 54.2% to 91.7% in the L2 group. The average accuracy scores of 70.05% and 72.53% while low, are above chance. The accuracy scores for Experiment 1b and the fillers are even lower than this overall average, which is artificially high due to the high accuracy scores beyond 90% for both groups for Experiment 3b. As the overall accuracy means are acceptable and the accuracy for Experiment 3b is especially high, I conclude that participants were indeed paying attention and no participant's data will be excluded from the self-paced reading data analysis based on his or her accuracy score. A between-groups ANOVA that compared the accuracy scores for items from Experiment 1b, Experiment 3b and the filler items revealed no main effect of Group ($F_1(1,62)=0.98$, $p=0.326$), no Group x Experiment interaction ($F_1(2,124)=1.58$, $p=0.21$), but a main effect of Experiment ($F_1(2,124)=98.43$, $p<0.001$). As is already visible from Table 4.9 above, the two participant groups did not differ statistically in their accuracy scores. The low accuracy scores of some participants in Experiment 1b likely reflect the challenging nature of the task. Leaving aside the filler questions, questions from Experiment 1b that tested the actual manipulation had significantly lower accuracy scores than questions from Experiment 3b that only tested the general comprehension of the sentence: $t(63)=9.42$, $p<0.001$.

Data cleaning procedure – self-paced reading data

The original dataset contained data from 32 participants in the L1 group and 32 participants in the L2 group resulting in 1536 trials with ten words per sentence.

In a first step, the data of two L2 participants was removed as they had shown very low accuracy for non-canonical word orders in general in the agent identification task and had reported not to know the object topicalization word order, which led to 0% accuracy in this condition (48 trials, 3.12%). Even if the agent identification task assessed conscious knowledge and the SPR task taps into unconscious knowledge, it would not be clear how these two participants interpreted the OVS sentences that they read. It would not be possible to tie possible changes in reading times patterns to an actual reanalysis or just a reading disruption caused by an unfamiliar structure.

In a next step all trials containing unknown words indicated in the vocabulary sheet by the L2 group were removed (44 trials, 2.86%). After this removal procedure, the dataset still contained 1444 trials. To prepare for the following data analysis, extreme values and outliers were removed for each segment separately as the segments were of different length and the last segment containing sentence wrap-up effects would have disproportionately affected the exclusion criteria, if one criterion had been applied to all segments equally. Extreme values were always identified visually by using a histogram, while outliers were identified based on a +/- 2.5 SD range around the participant mean per segment. A table with the cutoff points for extreme values and number of outliers removed for each segment can be found in Appendix B. No more than 51 data points were removed for any one segment, corresponding to 3.53% of all data points.

Main analysis

Table 4.10 below gives an overview of the average reading times per condition for both participant groups separately after applying the data cleaning procedure described above. As can be seen from the table, the L2 group had overall slower reading times than the L1 group. I therefore expect main effects of Group in all subsequent analyses.

A graphical depiction of the reading times across the entire sentence can also be found in Appendix B. Figures 4.10 and 4.11 display a focus on the region of manipulation (NP2 and main verb), the pre-manipulation region (auxiliary) and the spillover region (preposition).

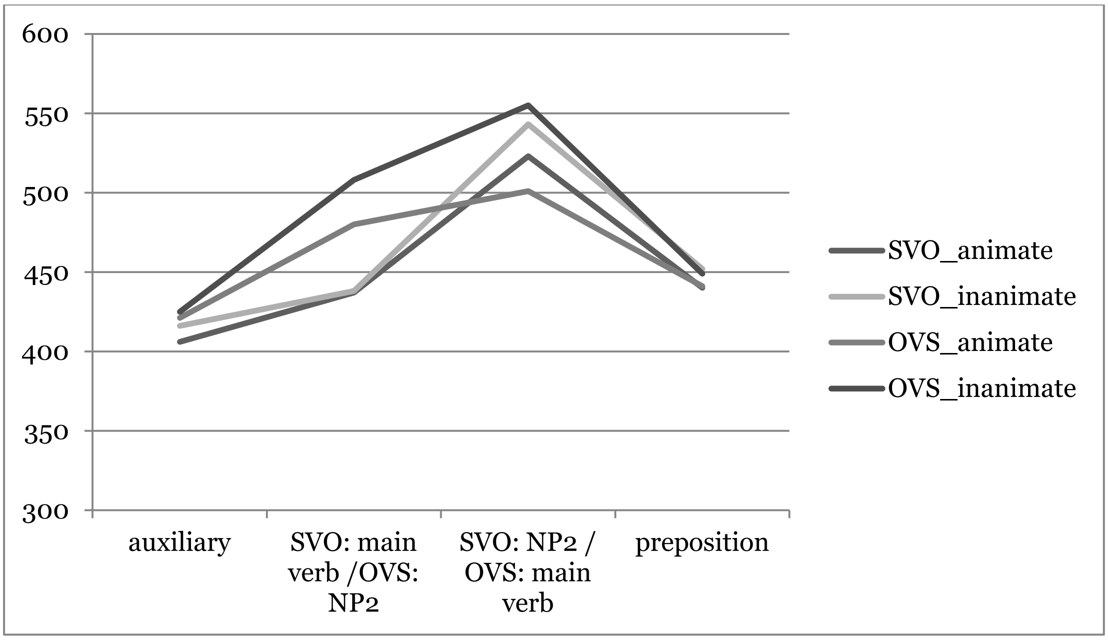

Figure 4.10 Mean reading times (in msecs) critical region; L1 group (Experiment 1b)

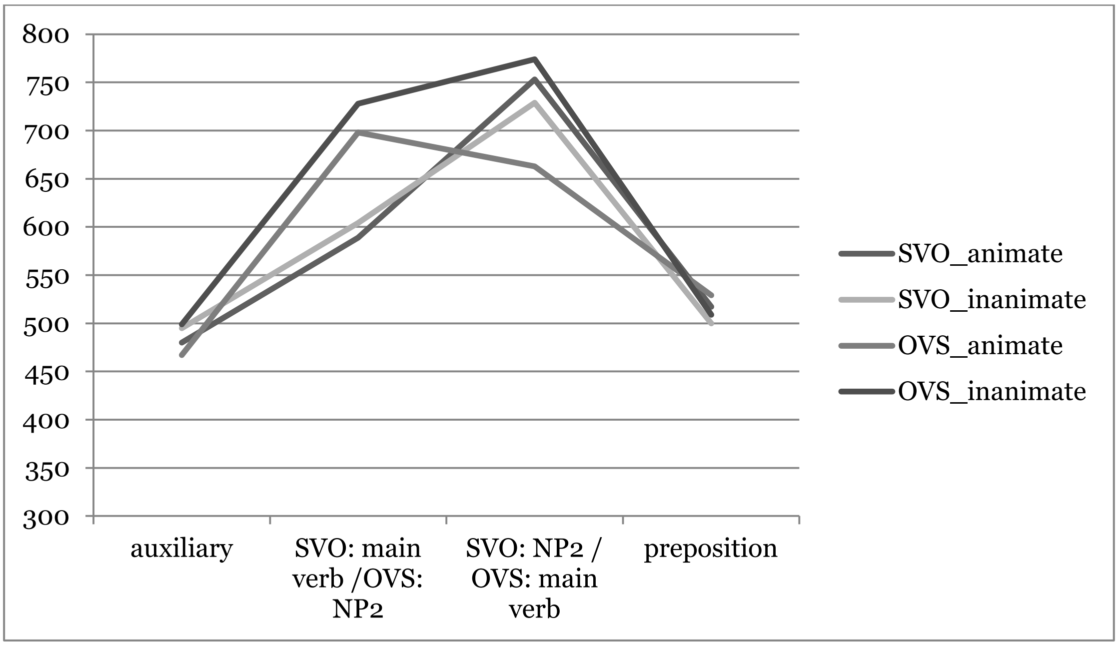

Figure 4.11 Mean reading times (in msecs) critical region; L2 group (Experiment 1b)

	NP1	relative pronoun	adverb	verb	adjective	auxiliary	main verb (SVO)/ NP2(OVS)	NP2(SVO)/ main verb (OVS)	prep	NP
L1 group (N=32)										
SVO_animate	621 (250)	427 (106)	406 (110)	403 (115)	663 (345)	406 (67)	437 (150)	523 (168)	440 (68)	1043 (758)
SVO_inanimate	596 (190)	415 (109)	409 (110)	396 (102)	643 (304)	416 (76)	438 (163)	543 (178)	452 (104)	1037 (635)
OVS_animate	603 (177)	420 (108)	425 (115)	409 (94)	634 (282)	421 (73)	480 (147)	501 (133)	441 (92)	1187 (1157)
OVS_inanimate	582 (200)	416 (116)	414 (123)	398 (103)	588 (266)	425 (77)	508 (191)	555 (172)	449 (74)	1324 (1276)
L2 group (N=30)										
SVO_animate	862 (311)	477 (116)	510 (159)	464 (113)	921 (466)	480 (108)	589 (221)	753 (276)	517 (121)	1269 (700)
SVO_inanimate	852 (322)	478 (136)	513 (131)	476 (120)	943 (494)	495 (146)	604 (235)	729 (353)	500 (107)	1284 (1050)
OVS_animate	868 (315)	476 (163)	521 (186)	468 (135)	894 (469)	467 (117)	698 (279)	663 (243)	529 (184)	1584 (1251)
OVS_inanimate	880 (371)	475 (161)	504 (140)	445 (106)	943 (559)	499 (142)	728 (344)	774 (364)	509 (118)	1702 (1326)

Table 4.10 Overview of reading times per segment per group, manipulation is shaded, SDs given in brackets (Experiment 1b)

If the position in which the verb and the NP2 occur within the sentence had no effect on the reading times, we would see a perfect X-shape in Figures 4.10 and 4.11 as the verbs were shorter than the nouns and therefore likely to be read more quickly than the nouns. The nouns on the other hand had been matched for length and frequency, so if animacy played no role, the reading times for the NPs should not differ. Visual inspection of the two figures above clearly shows that reading times were affected by verb position and partly also by NP2 animacy. While the reading times for the main verb in the SVO conditions differ very little, there is a very big discrepancy between the reading times for the main verb in the OVS conditions. In both groups, they correspond to the fastest reading time and the slowest reading time in the second segment of the manipulation. The difference was 54ms in the L1 group and 111ms in the L2 group. The NP2s also showed a difference in reading times, but the two groups seem to diverge here. The L1 group shows a comparatively big difference between the animate and the inanimate NP2 in the OVS condition, but not in the SVO condition. The L2 group shows a much smaller difference between animate and inanimate NP2 in either condition, despite the L2 group's overall greater variance. The above graphs also show that differences between the conditions before and after the manipulation are minimal.

The box-cox power transformation suggested a log-transformation for seven out of ten segments (a reciprocal square root transformation and a reciprocal (inverse) transformation in the remaining cases). In order to apply the same type of analysis to all of the data and due to the fact that a log-transformation was suggested for both segments that were manipulated, all ANOVAs were calculated on log-transformed reading times. I will give a short summary of the results of the ANOVAs run on the segments before the manipulation. For more detailed information on the results, the reader is referred to Appendix B. In all six segments, the main effect of Group was at least marginally significant due to the longer reading times in the L2 group. Only one other marginally significant effect was found: a main effect of Order in the F_2-ANOVA for the adjective that ended the relative sentence. This is likely a spurious effect as the participants at this point did not have evidence for a non-canonical word order. An effect of animacy could have been due to the variance of adjectives based on the preceding noun as some adjectives did not fit with inanimate NPs.

Except for this one effect, present only in one segment and in the by-item ANOVA, there was no influence of condition before the actual region of manipulation.

In order to be able to compare segments of similar size, I summed up the two regions of manipulation (segment 7 and 8) to one unit for analysis. I did this as only the nouns had been matched for length and I wanted to avoid effects based solely on the fact that the verbs were shorter than the NPs. Separate analyses of the two segments can again be found in Appendix B and show different effects in the two segments (in addition to the omnipresent main effect of Group): a main effect of Order in segment 7 and a main effect of Animacy and an Animacy x Order interaction in segment 8. After a visual inspection of the reading times of all participants, values above 6000ms were removed (1 data point, 0.07%). The box-cox power transformation suggested that the data be log-transformed, and outliers that were 2.5 SD above or below the subject mean were removed from the log-transformed data (33 data points, 2.28%)[6]. The between-groups ANOVA showed main effects of Group ($F_1(1,60)=16.79$, $p<0.001$, $F_2(1,23)=223.23$, $p<0.001$), Order ($F_1(1,60)=22.24$, $p<0.001$, $F_2(1,23)=14.76$, $p<0.001$), and Animacy ($F_1(1,60)=4.9$, $p=0.031$, $F_2(1,23)=2.44$, $p=0.13$), and an Animacy x Order interaction ($F_1(1,60)=8.21$, $p=0.007$, $F_2(1,23)=1.27$, $p=0.27$). As there was no interaction with Group, the data of the two groups was not analyzed separately.[7] Figure 12 below displays the raw reading times per condition. The main effect of Order is reflected in the faster reading times in the SVO condition. The main effect of Animacy is significant only in the by-subjects analysis, but is visible in the lower average reading times for animate subjects. The Animacy x Order interaction was also only significant in the by-subjects analysis and figure 12 suggests that this is due to the presence of an effect of animacy in the OVS condition, while it is absent in the SVO condition. Post hoc t-tests confirmed this impression: $t_1(61)=-3.15$, $p=0.0025$ for the OVS condition and $t_1(61)=-0.19$, $p=0.85$ for the SVO condition.

[6] Note that the number of data points removed is not identical to the number of data points removed for segment 7 and 8 reported in the appendix, since a different criterion was used for extreme value removal and outlier removal was repeated for this summed segment separately.
[7] The by-segment analyses showed a few marginally significant interactions with group. Separate analyses for each group can be found in Appendix B.

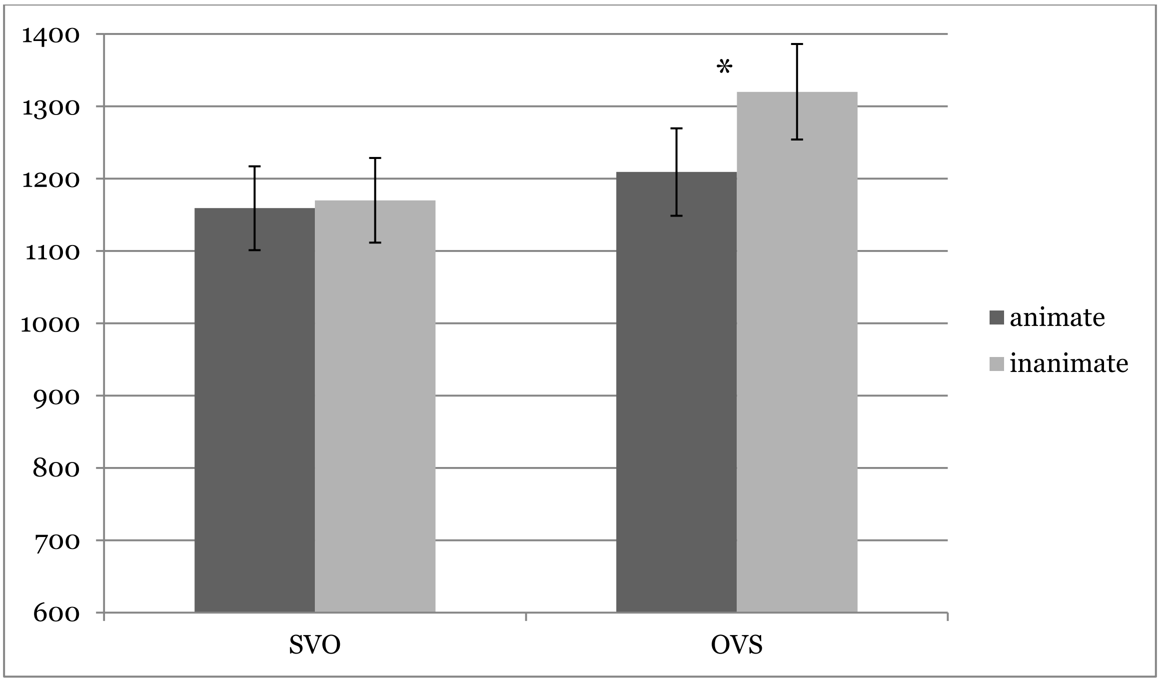

Figure 4.12 Reading times in the region of measurement per condition (Experiment 1b)

Next, I investigated the spillover region that was very short as it only consisted of a preposition. The between-groups ANOVA run on the log-transformed reading times only showed a significant main effect of Group ($F_1(1,60)=7.93$, $p=0.007$, $F_2(1,23)=75.4$, $p<0.001$). A Group x Animacy interaction failed to reach significance ($F_1(1,60)=2.82$, $p=0.098$, $F_2(1,23)=3.35$, $p=0.08$), all other $Fs<1$. Effects of the manipulation were therefore not carried over to the next, admittedly very short, segment.

The analysis of the sentence-final segment needs to be interpreted cautiously as it entails sentence wrap-up effects and might not reflect processes directly associated with the experimental manipulation. Apart from the main effect of Group ($F_1(1,60)=3.86$, $p=0.054$, $F_2(1,23)=105.84$, $p<0.001$), there was a main effect of Order ($F_1(1,60)=23.54$, $p<0.001$, $F_2(1,23)=6.4$, $p=0.02$). A look at table CV shows that this main effect is caused by the longer reading or wrap-up times for OVS sentences compared to SVO sentences in both participant groups.

Post-hoc analyses

Linear regression using L2 specific data such as self-rated proficiency, AoA, years learning Norwegian and stay in Norway as predictors and run on the same reading time differences as the previous analysis did not show any significant results (all $ps>0.25$). An overview of the results can be found in Appendix B.

Discussion

Comprehension questions

The accuracy data from the comprehension questions in this experiment suggest that L1 and L2 speakers had difficulties giving correct answers. The lower accuracy score compared to the questions from Experiment 3b could be due to the nature of the questions as they inquired the actual experimental manipulation. By the time the participants encountered the comprehension question, they had to keep a 10-word long sentence in memory which was taxing to working memory. In order to correctly answer the question, the participants also had to have reached an unambiguous, fully specified interpretation of the sentence as a 'good enough' representation featuring the action and the two NPs would often not be sufficient for a correct answer. Underspecified representations of the experimental sentence would not allow a correct answer in many cases due to the use of passive, little differences in plausibility etc. Despite those difficulties with regard to question structure, the L1 accuracy scores for the OVS condition fare well in comparison to previous studies. Kristensen, Engberg-Pedersen, & Poulsen (2014) reported 51.1% accuracy for Danish L1 speakers for context-free OVS sentences, Bader & Meng (1999) reported between 34% and 55% accuracy for object-first sentences for German L1 speakers. The accuracy scores of the L2 group are not surprising either. Jackson (2008) reported 60% and 78% for English native speakers on German object-first wh-questions and Gerth, Otto, Felser, & Nam (2015) reported between 32% and 51% accuracy for L2 speakers on object-first sentences. My two experimental groups lie within these scores, and any variation among the different scores can be explained by L2 proficiency, the individual structures investigated, the methodology used or the questions themselves.

Reading times

The pattern found in the reading times analysis corresponded roughly to prediction D that assumed an effect of animacy only in the non-canonical order. As SVO orders are the canonical word orders their processing required no reanalysis and proceeded effortlessly for both animacy settings. Although animate subjects had been rated as more plausible subjects in the SVO sentences used in the pretest, this difference in plausibility did not seem to affect the processing in the SVO condition. This could be explained by the fact that as SVO is the canonical word order processing speed was at its fastest in the region of manipulation and the participants did not experience a disruption of their processing by the difference in subject animacy.[8] Effects of the experimental factors had also been limited to the region of manipulation. Processing models differ with regard to their ability to explain this result and I will discuss them in turn. I will start with the syntax-first approach within the generative framework adopted by Åfarli & Eide (2003). For convenience, I repeat the syntax trees of Figures 4.1 and 4.2 at this point:

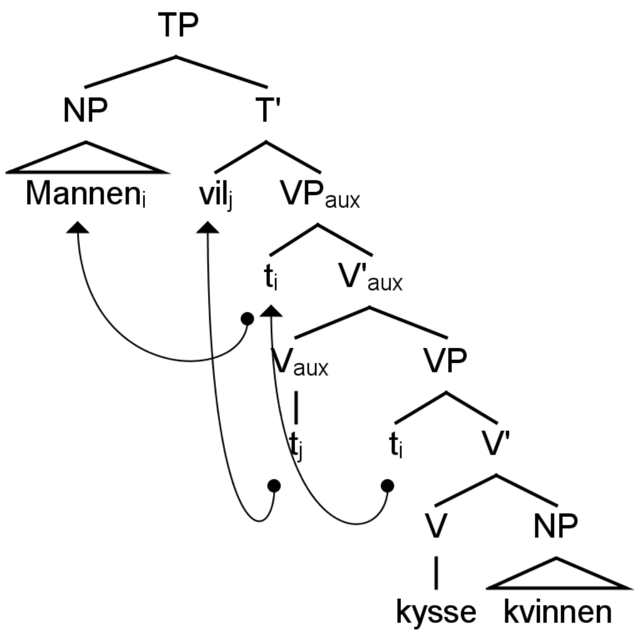

Figure 4.1: X-bar structure of a canonical Norwegian auxiliary sentence

[8] The by-group analysis reported in Appendix B shows a main effect of Animacy in the L1 group as this group shows faster reading times for animate subjects also in the SVO condition. The between-groups ANOVA masks this effect.
The separate L1 analysis showed a pattern that was in line with prediction C that assumed main effects for both factors, but no interaction. Subject animacy would therefore have an influence on native processing behavior even in canonical sentences. This finding is also in line with the higher plausibility for animate subjects in SVO that was found in the pretest that was conducted with native speakers.

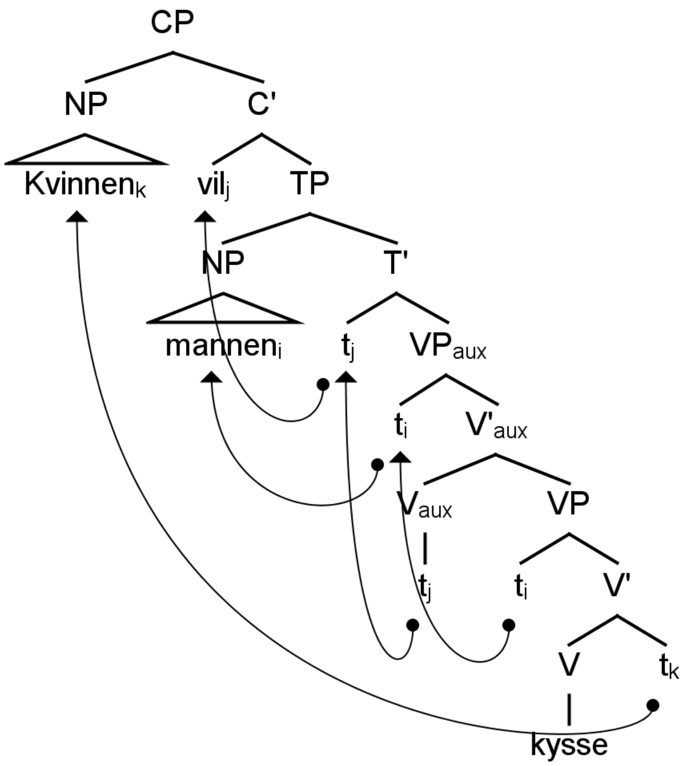

Figure 4.2: X-bar structure of a Norwegian auxiliary sentence involving a topicalized object

A syntax-first processing account would assume a massive reanalysis of the syntactic structure in OVS sentences as already pointed out during the first description of the syntactic trees in Figures 4.1 and 4.2. A number of filler-gap dependencies need to be established that are not present in the SVO sentence, among those the long dependency between the moved object and its gap-site in the lower VP. Except for the main verb, all other constituents (NP1, NP2, auxiliary) are affected by this reanalysis. The NP1 undergoes syntactic (from SpecTP to SpecCP) and thematic reanalysis (from agent to patient). The auxiliary has to occupy a position further up in the syntactic tree and leaves an additional trace. The NP2 is what signals the non-canonicity of the sentence and triggers reanalysis. This effortful syntactic reanalysis should cause higher reading times for OVS sentences in general according to strict syntax-first models. However, these cannot explain the effect of animacy found in the reading times for OVS sentences as the syntactic revisions needed remain the same independently of NP animacy. The thematic reanalysis that goes with the syntactic analysis can help to explain why OVS sentences with animate subjects show faster reanalysis times than those with inanimate subjects. In the thematic reanalysis of OVS sentences with animate subjects, the reanalysis changes the less favorable association of an inanimate NP with agency to the preferred association of an animate NP with agency. The opposite is true for OVS sentences with inanimate subjects that involve a reanalysis from a preferred to a dispreferred thematic role assignment. If we follow Bornkessel et al. (2003) in that thematic reanalysis processes can cause elevated reading times just as syntactic reanalysis can, then syntax-first models are able to explain the effect of animacy found in the OVS sentences as syntactic and thematic reanalysis cannot be separated in the experimental sentences used.

As there are no corpus studies investigating the occurrence of object topicalization in auxiliary sentences, Øvrelid's (2004) corpus study of OVS sentences in ambiguous NP1 V NP2 structures has to serve as an approximation for the discussion along the lines of models placing a stronger role on surface frequency. Øvrelid's study had found a 10% rate of occurrence for OVS structures, meaning that SVO orders should be processed a lot more quickly than OVS orders based on the more frequent order. Øvrelid's study has also found frequency differences with regard to NP animacy and object topicalizations in ambiguous sentences. All of the object topicalizations she found were of the type animate subject/inanimate object, similar to (77a). OVS orders were impossible whenever the two arguments were equal in animacy or when the subject was lower in animacy than the object (similar to 77b). This finding suggests that the availability of the movement of the object argument from VP to CP might be determined by the animacy of the arguments involved, which is not compatible with traditional generative frameworks. In ambiguous sentences, object topicalization is only available for subjects that are higher in animacy than the object and blocked in all other cases. While these restrictions based on animacy likely serve to avoid more ambiguity in NP1 V NP2 sentences, restrictions to the availability of object movement should not be necessary in unambiguous auxiliary sentences as syntax clearly assigns the roles of subject and object. If the conditions under which object topicalization is available are carried over from ambiguous to unambiguous sentences, OVS sentences like (77b) should have a surface frequency of 0. From a frequency perspective this would put them at a great disadvantage compared to OVS sentences like (77a) that are likely a lot less frequent than SVO sentences, but do still occur. Assuming that SVO sentences make up 90% of all auxiliary sentences, it is a bit surprising though that the OVS sentences with animate subjects do not show a bigger reading time disadvantage compared to the SVO condition, but are instead read almost as fast as the SVO sentences. In the absence of corpus data for object topicalization in auxiliary sentences, this explanation is speculative and needs additional support from said corpus studies.

The predictions of the Competition Model are based on the interaction of the cues main verb/NP2 order, subject animacy and NP1 animacy and the cue hierarchies in (83) and (84). The slowdown that participants experienced for inanimate subjects in the OVS condition compared to the OVS condition with animate subjects can be explained by competition between the main verb/NP2 order cue and the animacy cue. The word order cue has the less favorable OVS value in both conditions. However, in the OVS_inanimate condition, the NP1 is animate which corresponds to the favored value that NP1s are subjects and subjects are animate. At the point of disambiguation, two cues support an SVO interpretation with an animate subject, while only the word order cue supports the OVS condition. The two strongest cues as suggested by the hierarchy in (83) are in direct opposition creating a lot of competition between the two interpretations. Additionally, the favored interpretation of NP1=animate=subject has to be abandoned for the less favored NP1=animate=object and NP2=inanimate=subject. This is not the case in the OVS_animate condition in which the NP1 is inanimate and possibly less favored as subject. Here, only the preference for the NP1 to be the subject supports the SVO interpretation, while the word order cue and the animacy of the NP2 support an OVS interpretation. The two strongest cues, word order and animacy, converge in this condition, thereby reducing the competition compared to the OVS_inanimate condition. The interpretation of the subject also changes from the dispreferred inanimate NP to the preferred animate NP. If animacy is the strongest cue in the Norwegian cue hierarchy, it could speed up reanalysis to the point that it is almost cost-free and only slightly slower than the two SVO conditions.[9] The main problem is that the cue hierarchy suggested in (83) is based on the agent identification task in experiment 1b and not as rigorously tested as other cue hierarchies in the CM framework. Even if the order of the animacy and main verb/NP2 order cues were reversed, the convergence and competition in the two OVS conditions would remain the same.

Overall, a pure syntax-first account without reference to a possible thematic reanalysis would not be able to explain the results. Frequency-based processing models and the CM are able to explain the results, but require further support from corpus studies and rigorous agent identification studies to verify their assumptions with regard to the frequency of Norwegian object topicalization and the Norwegian cue hierarchy.

4.5 Discussion

Experiments 1a and 1b on object topicalization in Norwegian yielded comparable results in the agent identification and the self-paced reading task. Object topicalizations with inanimate subjects had higher error rates and higher reading times than object topicalizations with animate subjects. Native and non-native groups showed highly similar patterns in both tasks. I will now address the research questions that are repeated here from Section 4.1 and summarize the possible answers provided by the data.

Q1.1 Are Norwegian native speakers and advanced German L2 speakers of Norwegian able to correctly interpret sentences with topicalized objects in the absence of a supporting context based only on the order of the main verb and the second NP?

The results of the offline task suggest that about a third of all participants (native and non-native) can reliably use the main verb/NP2 word order cue to identify the agent in an OVS sentence. The remaining participants used the word order cue successfully only in some of the OVS sentences. In the self-paced reading task

[9] A CM approach could also explain the effect of animacy in the SVO condition found the in L1-only data analysis. If animacy is the strongest cue in the hierarchy and stronger than the word order cue, it would also be beneficial in SVO sentences with animate subjects as the presence of an animate NP in the object position would create competition.

participants showed slower reading times in the OVS condition compared to the SVO condition suggesting that they were all sensitive to the change to the sentence structure. It is not clear, however, whether this sensitivity had the same consequences for all participants. Some might have pursued a successful analysis, some might have resorted to shallow parsing, and even a breakdown of parsing is possible. Elevating the number of comprehension questions in order to use them as a dependent variable could allow more insights into the parsing outcomes of individual subgroups of participants. The accuracy rate of the agent identification task at 65% is quite low for an untimed, offline task and shows high individual variation suggesting that identifying the agent in a Norwegian OVS sentence is a difficult task and not as straightforward as suggested by descriptive grammar. Nevertheless, about the same amount of participants from the L1 and the L2 group were able to reliably identify a topicalized object despite the absence of a supporting context, no case marking and no verb agreement, the disambiguating cues that are usually used in subject/object ambiguity studies.

Q1.2 What role do general ordering principles such as animate > inanimate, agent > patient play in the interpretation of object-topicalized sentences? Are sentences harder to interpret if they violate these principles? Is the disruption equally strong for all principles or do some principles have a stronger influence on the interpretation than others?

The animate > inanimate principle seems to be of great importance at least in the native processing of Norwegian object topicalizations, but not in the traditional way that animate NPs preferably precede inanimate NPs. In Norwegian, the application of this principle is closely tied to subjecthood. Animate NPs mostly precede inanimate NPs, if they are the subject of the sentence supporting the strong association of animacy and subjecthood. Topicalization of animate objects is not possible in ambiguous sentences (NP1 V NP2). The topicalized object would be equal or higher in animacy than the subject. The preference for subjects to be animate and sentence-initial would likely lead to a misinterpretation of the topicalized animate object as the subject of the sentence.

The online and offline data of the L1 group showed evidence that the blocking of object topicalization for animate NPs also applies to unambiguous auxiliary sentences in some speakers. OVS sentences that contained subject-object animacy combinations that would disallow object topicalization in ambiguous sentences showed elevated error rates for the L1 group in the agent identification task. In the self-paced reading data the OVS sentences differ with regard to their violations of ordering principles. The OVS_animate condition violates the animate > inanimate principle, but adheres to the principle that topicalized objects should be lower in animacy than the corresponding subjects. The OVS_inanimate condition follows the animate > inanimate principle, but violates the animacy principle for object topicalization as the topicalized object is animate with an inanimate subject following. The reading time data showed that the latter condition caused a stronger disruption of the parsing process, suggesting that the animate > inanimate principle is less influential than the animacy principle for topicalized objects and the association of animacy and agency.

The ordering principle of agent > patient is confounded with the word order manipulation in the online task, as all subjects were the agents of the sentences. Both SVO conditions therefore follow the agent > patient principle, while the two OVS conditions violate it. When comparing the reading times of the four conditions, there is a big difference between the two OVS conditions, but a much smaller difference between the OVS_animate condition and the two SVO conditions. If only the violation of the agent > patient principle had an influence on the reading times, the OVS_animate condition should have been read much more slowly than the SVO conditions with a smaller or no difference between the two OVS conditions. It seems that the agent > patient principle is also modulated by NP animacy. The same trend was also found in the L1 error rates in the agent identification task which showed generally elevated error rates for the OVS conditions, but also an influence of NP animacy. The two ordering principles in isolation are not able to able to explain the L1 results in the offline and online task.

The L2 group showed no influence of NP animacy in the offline task with just a general effect of word order and possibly the associated agent > patient principle. The application of ordering principles in the online data seems identical to the native speakers in the OVS condition. This is to say there was no influence of the animate > inanimate principle as discussed above, some influence of the agent > patient principle with slower reading times in the OVS conditions, and an overall stronger influence of a principle that associates animacy and subjecthood.

Q1.3 Is both syntactic (i.e. phrase structure) and lexical-semantic information (i.e. NP animacy) considered during the processing of object-topicalized sentences? Do L1 and L2 speakers differ in their use of the two sources of information?

The results of the self-paced reading task suggest that syntactic and lexical-semantic information are indeed both considered during the processing of object topicalization sentences. There was no difference between the L1 and the L2 group in their use of both information sources in the OVS condition. Both groups showed signs of reanalysis based on the syntactic information, as reading times were elevated at the region of manipulation for OVS conditions compared to SVO conditions. Reading speed clearly benefited from a reanalysis towards the preferred association of animacy and subjecthood, as was visible by a smaller increase in reading times in the OVS_animate condition compared to the increase in reading times in the OVS_inanimate condition, which required a reanalysis towards a less favorable inanimate subject. L2 processing accounts like the SSH propose that the processing of non-local dependencies in L2 speakers is based more on lexical-semantic information than on syntactic information compared to the more syntax-driven processing of L1 speakers. These accounts usually predict different results for native and non-native speaker groups in processing tasks. The data from experiment 1b did not show a difference between L1 and L2 processing. This might seem to be less compatible with accounts like the SSH, but it is not impossible to explain the result. Differences between L1 and L2 processing are assumed to occur, because L2 speakers tend to underuse the syntax-driven parsing route. Now, if the L1 group also underuses this parsing route and is guided heavily by lexical-semantic information like animacy, no difference between L1 and L2 processing would be predicted even by the SSH. Whether the L1 group in this task resorted to animacy-driven processing, either because it is their usual processing routine in the absence of reliable syntactic information or because the challenging nature of the task pushed them towards ‘good enough’ processing, is a topic of future research and cannot be evaluated based on the present data.

5 Study 2: Object order in ditransitive sentences in German

Similar to the previous experiment reported in Chapter 4, this experiment investigated object movement. But in this experiment we change the language to German and are no longer looking at object topicalizations, i.e. movement to the front of the sentence, but instead at scrambling, i.e. the movement of objects within the German midfield (the section following the finite verb or complementizer and preceding a potential non-finite verb or particle).

As the literature review in Sections 3.2.2 and 3.2.3 suggested, this structure is fairly well documented in L1 processing studies using main and embedded clauses and a variety of experimental techniques (Bader & Meng, 1999; Bornkessel et al., 2002; Pappert et al., 2007; Rösler et al., 1998; Schlesewsky et al., 2003). Corpus studies have established its frequency of occurrence (Heylen, 2005; Heylen & Speelman, 2003; Kempen & Harbusch, 2005), and Keller (2000) is a thorough, systematic investigation of its gradience in acceptability. There are lots of L2 studies on the English dative alternation, mainly production and rating tasks (De Cuypere et al., 2014; Marefat, 2005; Park, 2011). Many of the L2 rating studies have serious design flaws, though, undermining their conclusions that L2 speakers are sensitive to the factors that influence the occurrence of either rendition of the dative alternation. In addition, in these studies the language combinations are often such that the L1 is a non-scrambling language (often English), but see Hopp (2005) for a comparison of two L2 groups, one with an L1 that has scrambling (Japanese), and the other with an L1 that does not (English). Studies on scrambling in L2 German are rare and often restricted to ratings like Hopp (2005) and the study in this thesis compares ratings with online processing in the same participants. As both scrambled and non-scrambled word orders are grammatically correct, Study 2 also seeks to investigate whether native and non-native speakers make the same distinctions with regard to gradient acceptability. For this purpose, an acceptability judgment task and an SPR task were conducted with German native speakers and L2 speakers of German from a variety of Slavic L1 backgrounds. Slavic languages have free word order and the L2 speakers in this experiment come equipped with preferences from their L1s which will be addressed in Section 5.1. The acceptability judgment data therefore adds a new group of L2 speakers to the existing offline findings. The two tasks reported in this experiment did not make use of the full range of object ordering possibilities in German, but focused solely on the contrast of DAT > ACC and ACC > DAT orders with the subject always in its canonical position.

The following Section 5.1 provides an introduction to the structure of German main and embedded clauses with ditransitive verbs, involving an introduction to the terms prefield and midfield and to some select syntactic approaches. It also reviews hierarchy principles that have been investigated in the ordering of arguments in German and gives an overview of ordering preferences in the various Slavic L1s present in the L2 group. The research questions for this topic can be found at the end of this section. Section 5.2 reports the results of the acceptability rating task and Section 5.3 the results of the self-paced reading task. Section 5.4 summarizes the results of both experiments, addresses the research questions and relates the experimental results to the theoretical background. Finally, Section 5.5 concludes part I of this thesis with an intermediate summary on the findings regarding objects in non-canonical positions.

5.1 Background: Object order in ditransitive sentences in German and Slavic languages

This experiment investigated the processing of two possible orders of objects in ditransitive structures in the German midfield. The section preceding the finite verb, known as the prefield, must be filled with exactly on constituent. In contrast, the midfield allows the presence of several constituent and their order is flexible. In sentences with a ditransitive verb, all three arguments can occur in the midfield and all six possible orders are equally grammatical. Table 5.1 exemplifies the German field structure for a main clause and a subordinate clause. In both cases, the subject *der Vater* (the father), the indirect object *dem Sohn* (the son) and the direct object *einen Ball* (a ball) are found in the midfield. The main difference between the two sentence types is the position of the finite verb (printed in bold in Table 5.1). It appears in the first bracket in the main clause and in the second bracket in the subordinate clause.

		prefield (Vorfeld)	1st bracket (Klammer)	midfield (Mittelfeld)	2nd bracket (Klammer)	back field (Nachfeld)
Main clause		Gestern	**hat**	der Vater dem Sohn einen Ball	geschenkt,	weil er Geburtstag hatte.
	Yesterday, the father has given the son a ball, because it was his birthday					
Subordinate clause	Er sagte,		dass	der Vater dem Sohn gestern einen Ball	geschenkt **hat**,	weil er Geburtstag hatte.
	He said that the father has given the son a ball yesterday, because it was his birthday.					

Table 5.1 Comparison of German main clause and subordinate clause adapted from Heylen and Speelman (2003)

Although the argument order in the midfield is free, it is influenced by various linearization factors, such as case, pronominality or animacy. These factors are either thought of as being in competition with each other (Siewierska, 1993; Primus, 1994), or construed as a subhierarchy in an optimality theoretical (OT) framework (Müller, 1999:795). Despite the differences between the explanation of whether these factors interact or not and how they fit into certain theoretical frameworks, there is general agreement as to the direction of some factors. The list of ordering preferences in Table 5.2 is not supposed to reflect any hierarchy between the different preferences.

Ordering Factor	**Direction of preference**
Nominative	Nominative > Non-nominative
Pronominality	Pronoun > Non-pronoun
Dative	Dative > Accusative
Definiteness	Definite > Indefinite
Animacy	Animate > Inanimate
Focus	Non-focus > Focus

Table 5.2 List of ordering principles found to influence word order in German

These linearization factors have mainly been established by corpus studies that often considered only single factors in isolation (e.g. Animacy by or three separate analyses for Dative, Animacy and Definiteness by Pappert et al., 2007) or at most two factors and their interaction (e.g. Case and Pronominality by Weber & Müller, 2004). The factors Nominative and Dative in Table 5.2 can be summarized under one bigger factor Case and when combining the two factors, the canonical order NOM > DAT > ACC emerges. There is evidence in the literature, however, that the two ordering principles operate independently, as the violation of Nominative causes a bigger disruption in processing than the violation of Dative (e.g. Rösler et al., 1998). In his thesis, Keller (2000) systematically investigated ordering constraints and their influence on the acceptability of a variety of different word orders. For ditransitive subordinate clauses, he specifically looked at the following three constraints: Nominative, Pronominality and Dative. Keller found that none of these constraints is absolute, meaning a violation does not result in extremely low acceptability ratings. The constraints Pronominality and Nominative turned out to have similar strength, while Dative was much weaker than the other two.

5.1.1 Corpus studies on the application of linearization principles in German ditransitive sentences

Corpus studies are a popular and useful tool in the study of order variations and the different principles that possibly govern the use of a specific order. However, the number of hits in corpora for ditransitive sentences in general is already small and the variation in the form of the constituents (pronominal vs. nominal) further complicates the investigation of specific combinations, e.g. ditransitive sentences with all nominal constituents. Many of the corpus studies reported below operate with frequencies of occurrence well below 50, sometimes as low as single-digit numbers for certain non-canonical orders.

Extending Keller's work, Kempen & Harbusch (2005) predicted that these ordering constraints should affect language production. Orders that agree with constraints would be more frequent than orders that violate constraints. They assumed a direct connection between the frequency of an order and its acceptability. In their study of the NEGRA-II (http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/negra-corpus.html) and the TIGER (http://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger.html) written corpora and the VERBMOBIL spoken corpus (http://verbmobil.dfki.de/) investigating German transitive and intransitive sentences, the authors found a discrepancy between the acceptability ratings reported in Keller (2000) and actual corpus frequencies of the same constructions. Orders that still received an average rating were never attested in their corpus and all orders found in the corpus were rated as highly acceptable in offline ratings. Kempen & Harbusch therefore suggest a production threshold in the grammaticality continuum.

Structures with grammaticality ratings above this threshold would be produced with medium to high frequencies, structures around the threshold would be perceived as marked and produced with very low frequencies possibly unattested in corpora. Any structures with grammaticality ratings below the threshold would be considered ungrammatical and their occurrence would either constitute an unintentional error caused by, for example, a false start in language production or an intentional distortion of the output.

Based on frequencies from the three abovementioned corpora, Kempen & Harbusch suggest a fixed word order NOM > ACC > DAT, if all three constituents are pronominal. If all constituents are full NPs, the picture is different. The orders NOM > DAT > ACC and NOM > ACC > DAT received comparable ratings in Keller's (2000) data, but the numerically predominant order found in the written and the spoken corpora combined is NOM > DAT > ACC with 54 occurrences against five for full subject NPs, and 17 occurrences against five for pronominal subjects. Non-pronominal object preposing in the midfield is rare and mainly caused by lexical-semantic properties of the verb, e.g. experiencer object verbs (DAT > NOM order) or certain ditransitive verbs (ACC > DAT order). Kempen & Harbusch (2004b, 2005) posit a *production-based linearization rule* that assigns a primary position to each argument NP (e.g. pronominal NPs have a more fronted position than full NPs) and a secondary, more anterior position for full NPs (see Figure 5.1.).

	NOMpro – ACCpro – DATpro – NOM – DAT – ACC		
Clause type	Transitive	Intransitive	Ditransitive

Figure 5.1 Production-based linearization rule adapted from Kempen & Harbusch (2005)

The movement of a full NP into its secondary position indicated by the arrows in Figure 5.1 can be triggered by conceptual factors like animacy, definiteness or referential ease (Kempen & Harbusch, 2004a, 2004c). The linearization rules represent structures that receive high grammaticality ratings, and the authors suggest that all the orders that can be derived from it are unmarked orders. Marked orders that are not part of the linearization rule are the result of the influence of strong conceptual factors like topic/focus orders. According to Kempen and Harbusch, the gradience found in acceptability judgments reflects the number and severity of violations of the production-based linearization rule in Figure 5.1. The authors also suggest that sentence production should be aimed at keeping the number of violations minimal. While these suggestions might be suited to explain the corpus data, they are at odds with the experimental data that did not find a gradient increase of processing difficulty with decreasing acceptability (e.g. Rösler et al., 1998).

Kempen & Harbusch (2004a) investigated whether the influence of animacy on constituent order is an indirect one or a direct one. The authors define an indirect influence of animacy on ordering preferences because of the association of animacy with agency and that of agency with subjecthood. Agents are preferably animate and subjects, therefore animate NPs tend to precede inanimate ones that are preferably patients and objects. This is similar to Primus' (1998) argument against an independent influence of animacy on ordering due to the tendency of agents and recipients to be animate, and therefore to precede inanimates based on her thematic hierarchy (see Section 3.1). One main difference is that Kempen and Harbusch also derive the S > IO ordering principle from this indirect influence of animacy. If we follow Primus in that recipients also are preferably animate, the preferred S > IO would not emerge, if based solely on the preferred animacy of the NPs involved. A direct influence of animacy according to Kempen and Harbusch assumes that conceptualization, role assignment and attachment to the surface structure take places earlier for animate NPs than for inanimate NPs. The authors cite some experimental production data that can support both hypotheses regarding the influence of animacy. Their own study, however, is a corpus study, although especially the direct influence hypothesis makes predictions on processing that cannot possibly be addressed by a corpus study. The corpus study itself also was problematic as there were only two subject/object combinations that yielded enough instances of a dispreferred order (according to syntactic ordering factors) to warrant an analysis. Both combinations (nominal subject and pronominal DO, nominal subject and nominal IO) showed effects of animacy on constituent order. Constituents that should have been placed in the second position were fronted more often when they were animate. Animacy did not completely overrule ordering preferences such as pronominal > nominal, but it attenuated their effects. Kempen and Harbusch interpret their result as a confirmation for a direct influence of animacy. Given that the direct influence hypothesis involves a prediction with regard to the timing in speech production, Kempen and Harbusch's data does not allow any conclusions with regard to timing or processing as it is corpus data and in addition a fairly small dataset.

Pappert, Schließer, Janssen, & Pechmann (2007) used the same written corpora as Kempen & Harbusch and additional completion questionnaires and self-paced reading experiments to study the influence of three constraints on the frequency and processing ease of word order variants of double object sentences. The constraints in question were the Dative constraint (DAT > ACC), the Animacy constraint (animate > inanimate) and the Definiteness constraint (definite > indefinite). The authors investigated ditransitive sentences with the subject in the prefield position like (91) below.

(91) Der Mann wird dem Kind den Brief geben.
the$_{NOM}$ man will the$_{DAT}$ child the$_{ACC}$ letter give.
'The man will give the letter to the child.'

This sentence type was chosen over an equivalent subordinate clause as it was more frequent in the corpora investigated. In the self-paced reading task, the sentence-final position of the subcategorizing verb delayed the assignment of the number and type of objects until the end of the sentence, when verb was reached and all objects had been encountered. The results for the Dative constraint showed a clear application in all three measures used. DAT > ACC orders occurred more frequently in the corpus, dative-first sentences were far more likely to be completed as a double object structure than accusative-first sentences, and the SPR experiment showed higher error rates and longer reading times for ACC > DAT sentences. The Animacy constraint was found to be strongly associated with the Dative constraint as dative objects were usually animate and accusative objects usually inanimate, both in the corpus and the completion data. Of the eight possible combinations of animacy and case, the order animate dative > inanimate accusative was the most frequent order in the corpus and the completion data. The strong association between case and animacy was explained by typical thematic roles as animate non-agents are typically recipients, while inanimate non-agents are typically patients. There was no SPR experiment to test this constraint. Evidence for the Definiteness constraint was only found in the corpus data, but neither in the completion data nor the self-paced reading data showed an effect of definiteness. In the corpus data, dative objects were mostly definite, while accusative objects tended to be more frequently indefinite than definite. Overall, the order definite dative > indefinite accusative occurred nearly twice as often as the next most frequent order definite dative > definite accusative suggesting an association between Case and the Definiteness constraint. The only tendency present in the completion data was a preference for exactly one indefinite object per sentence and longer reading times if both the Dative and the Definiteness constraint were violated by a sentence. The authors suggest that this could be due to the absence of a context in their experiments, while context is present in the corpus.

Kurz (2000) also used the NEGRA corpus to examine ordering hypotheses derived from Hawkins' Early Immediate Constituents principle (Hawkins, 1994). This principle assumes that orders are preferred that allow a faster assessment of all constituents in a sentence. Long constituents for example should preferably be in a late position as their early occurrence causes a delay in constituent identification as the long constituent first needs to be fully processed before the following constituent can be integrated. The parameters investigated by Kurz were length of the NP and definiteness. Three transitive and three ditransitive verbs were chosen for the analysis based on their frequency in the corpus. As Kurz only included sentences with nominal constituents in the midfield, the number of exemplars found in the corpus that fit this criterion was fairly limited as often one constituent was pronominal. Based on the order distributions that she found, Kurz advocates a verb-based determination of basic argument order and not a general basic argument order. Two transitive verbs (*gelingen* 'succeed' and *zur Verfügung stehen* 'to be available') showed a clear DAT > NOM preference, while *helfen* 'to help' had a clear NOM > DAT preference. For the ditransitive verbs, *geben* 'give' and *zur Verfügung stellen* 'to make available' showed a strong DAT > ACC preference, while the verb *vorstellen* 'to present' showed no preference as both argument orders were equally frequent. Kurz found weak predictive power of the EIC for rearranged orders and differences in length alone were not enough to explain rearranged orders. However, definiteness that is assumed to be epiphenomenal by the EIC seemed to play a bigger role with regard to ordering preferences. In basic word order, definite-definite and definite-indefinite sequences were the most frequent, while the predominant pattern in rearranged orders was the definite-definite sequence.

Another study using the NEGRA-corpus by Heylen & Speelman (2003) investigated the influence of the following factors on the ordering of the arguments: sentence type (main vs. subordinate clause), grammatical function and length of the arguments, animacy and given/new status. The main goal of this study was to increase the number of factors in order to achieve a model that is closer to actual language use. The authors investigated transitive structures with a subject NP and an object as personal pronoun such as *Ein paar Tage später nahm ihn* (object) *der SED-Chef der Uni* (subject) *beiseite* ('A few days later the university's SED-chief took him aside'). Ditransitive sentences were not analyzed because typical of most corpus studies on this topic, there were too few examples in the corpus. In this study the pronominal object was found to precede the nominal subject in 89% of all cases and clause type was found to significantly affect this percentage as the pronominal object was placed before the nominal subject 95.9% of the time in main clauses and 75.7% of the time in subordinate clauses. Accusative objects were overall more frequent than dative objects in the transitive sentences, but order preferences did not change with the type of the object. As the pronominal object always consisted of only one syllable, the difference in syllable number between subject and object was always an increase in the length of the subject. The predominance of the pronominal object-first order increased from 76.6% to 95.9% with increasing length of the subject. Givenness was assessed on a scale that reflected the mental accessibility of subject and object and ranged from 'to be created' to 'accessible in near context'. The more given a subject was, the more likely it was to occur before the pronominal object. Animate subjects also were more likely to occur before the pronominal object, but neither of these factors contributed to an occurrence of the nominal subject > pronominal object order beyond 25%. In their multivariate analysis, the ranking of factors from most influential to least influential was: clause type, given/new status, animacy and length difference, with grammatical status having no influence at all. In fact, the only time that there was a sort of 'free alternation' between the two orders with a smaller difference between subject-first and object-first orders was in subordinate clauses when the length difference between nominal subject and pronominal object was zero. In that case, the split was 58.3% for the object-first and 41.7% for the subject-first order.

Factor	**Direction**	**Study**
Nominative	nominative > non-nominative	strong constraint (Keller, 2000)
Pronominality	pronoun > non-pronoun	strong constraint (Keller, 2000) stronger in main clauses than in subordinate clauses (Heylen & Speelman, 2003)
Dative	dative > accusative	weaker constraint, little impact on acceptability ratings (Keller, 2000) violated if objects are pronominal, followed if all three constituents are nominal (Kempen & Harbusch, 2005) DAT > ACC order more frequent, more double object completions and faster reading times with higher accuracy in SPR (Pappert et al., 2007)
Definiteness	definite > indefinite	only found in the corpus data, not in completion or SPR task, associated with Dative constraint (Pappert et al., 2007) definite > definite most frequent in rearranged orders, definite > indefinite mainly in basic orders (Kurz, 2000)
Animacy	animate > inanimate	weaker constraint, more fronting of animate constituent (Kempen & Harbusch, 2004a) strong association with the Dative constraint (Pappert et al., 2007) animate nominal subjects are more likely fronted than inanimate nominal subjects (Heylen & Speelman, 2003)
Givenness	given > new	weaker constraint, given nominal subject more likely to occur before pronominal objects (Heylen & Speelman, 2003)
Length	short > long	weak constraint, cannot explain rearranged orders (Kurz, 2000) increasing length difference leads to more fronted pronominal objects (Heylen & Speelman, 2003)

Table 5.3 Overview of results of corpus studies with regard to ordering constraints

Table 5.3 summarizes the results of the previously reported corpus studies. It becomes evident that there are two strong constraints (Nominative and Pronominality) that dominate the other, weaker constraints. The Dative constraint seems to be the more powerful of the weaker constraints, as other constraints (Animacy, Definiteness) seem to be associated with it and exert an even weaker influence if considered independently. It has to be remembered, though, that the corpus studies reviewed often only looked at a subset of ordering constraints and for different constructions (transitives, ditransitives). And while some differences emerge in the degree of ordering preferences, e.g. between main and subordinate clauses, the overall ordering preferences are those presented in the second column in Table 5.3. It is clear, however, that ordering preferences are not guided by just one single principle, such as object length, but instead depend on the complex interaction of many difference principles with different strengths.

For Study 2 of this thesis I chose to investigate the application of the Dative constraint in L1 and L2 speakers in order to test a weak principle that might be challenging for L2 speakers as it depends on the correct processing of case marking compared to, for example, the Animacy principle that requires lexical knowledge. The Dative constraint was chosen as other principles are associated with it (Animacy, Definiteness) and it is less dependent on context for its correct applications than the Focus or the Givenness principle. The experiments were aimed to investigate the small differences in acceptability and processing that were found in L1 speakers when the Dative constraint was violated (Keller, 2000; Rösler et al., 1998) and not the big differences in acceptability that occur whenever the Nominative principle or the Pronominality principle are violated (Heylen & Speelman, 2003; Keller, 2000; Rösler et al., 1998). So far, there are no studies on the sensitivity of L2 speakers to gradient acceptability and how it might be reflected in their sentence processing.

Movement vs. base generation

From the corpus studies presented above, it has become evident that German allows different orders of the constituents in a sentence and that these orderings are not random, but are instead influenced by a variety of factors. The syntactic representation of these different argument orders can either involve some form of a movement approach (the one adopted in this thesis) or a base-generation approach. The movement account assumes one basic argument order in ditransitive sentences from which all other possible orders are derived either through A- or A-bar movement (depending on the author). In most accounts of German the basic order is the DAT>ACC order (see Georgala, 2011b for a discussion and Müller, 1999:802 for an opposite view). The competing base generation account assumes no basic word order, instead the various orders are all results of base generation (Kiss, 1994). There is not one commonly agreed upon syntactic representation of the verb and its two arguments in the movement account. Suggestions contain small clauses or VP-shells (Larson, 1988), branching can be strictly binary or have a flat structure with several nodes branching out (Fanselow, 1990), but they all have in common that the moved constituent leaves a trace.

For expository purposes, I adopt Larson's VP-shell analysis and IO > DO as the basic word order in German ditransitive sentences as seen in Figure 5.2. A base-generated representation would not assume any traces.

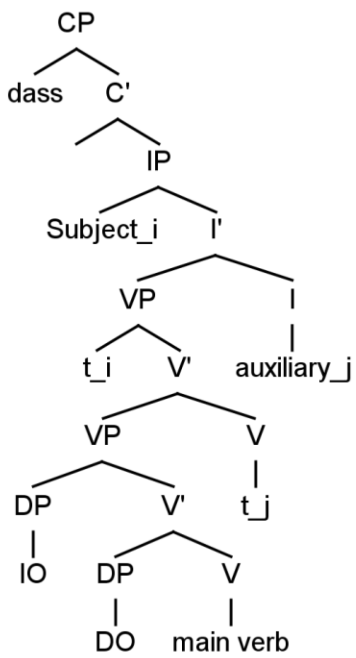

Figure 5.2 Syntactic tree representation of IO > DO order using a VP-shell structure

The DO > IO (or ACC > DAT) order is often used in German whenever the indirect object (IO) is focused and receives additional stress. This corresponds to the crosslinguistically frequently occurring ordering strategy of Theme > Rheme or Given > New. In a derivational approach this order is achieved via the movement of the DO up in the syntax tree leaving behind a trace in its original position. This additional trace that needs to be matched to its antecedent during processing should lead to higher processing costs compared to the unmarked IO > DO order. In a base-generation approach, the DO and IO are directly inserted

in their respective positions also in the DO > IO order without any traces. In the base-generation approach both orders are equally complex with regard to processing difficulty as no filler-gap dependencies need to be processed.
In addition to the two argument orders already mentioned, some verbs have a third option with the indirect object as a PP.

Double Object Construction (DOC): IO > DO
(92a) Ich sende dem Lehrer den Brief.
I send the$_{\mathrm{DAT}}$ teacher the$_{\mathrm{ACC}}$ letter
'I send the teacher the letter.'

DOC: DO > IO
(92b) Ich sende den Brief dem Lehrer.
I send the$_{\mathrm{ACC}}$ letter the$_{\mathrm{DAT}}$ teacher

Prepositional Dative Construction (PDC): DO > IO
(PP with accusative case marking)
(92c) Ich sende den Brief an den Lehrer.
I send the$_{\mathrm{ACC}}$ letter to the$_{\mathrm{DAT}}$ teacher
'I send the letter to the teacher.'

This third option is comparable to the well-studied Dative alternation in English that alternates between a double object construction (DOC) with the order IO > DO and a Prepositional Dative construction (PDC) in which the IO is rendered as a PP.

In addition to the verb-specific ordering preferences suggested by Kurz above in her corpus study there are attempts in the literature to characterize the two different German double object orders in (92a) and (92b) in similar terms as the English dative alternation. These characterizations are also based on verb-specific preferences and often focus on the thematic role of the IO, i.e. whether it expresses a benefactive or a goal. In English, a DOC is used preferably with benefactive IOs, whereas a PDC is used when the IO expresses a goal. Georgala (2011a, 2011b) goes beyond this dual distinction and suggests a tripartite classification of German DOCs based on Pylkkänen's (2002) theory of high and low applicatives.[10] This theory does, however, not seek to explain the German prepositional dative.

[10] Pylkkänen proposes two types of applicatives that express different relations between the arguments and the verb. High applicatives describe a relation between the individual and the event in the VP: [$_{\mathrm{VoiceP}}$ DP$_{\mathrm{AGENT}}$ [$_{\mathrm{Voice'}}$ Voice [$_{\mathrm{ApplP}}$ DP$_{\mathrm{BNF/LOC/INSTR...}}$ [$_{\mathrm{Appl'}}$ **Appl** [$_{\mathrm{VP}}$ V DP]]]]]. Low applicatives on the other hand describe a dynamic transfer of possesion between the IO and the DO: [$_{\mathrm{VoiceP}}$ DP$_{\mathrm{AGENT}}$ [$_{\mathrm{Voice'}}$ Voice [$_{\mathrm{VP}}$ V [$_{\mathrm{ApplP}}$ DP$_{\mathrm{GOAL/SOURCE}}$ [$_{\mathrm{Appl'}}$ **Appl** DP]]]]]. Note that the thematic

One group consists of the small and unproductive group of verbs with "low" datives that shows the basic word order of DO > IO and the dative as an oblique argument, such as (93) taken from Georgala (2011a). Verbs of this type were not chosen for the following two experiments.

(93) dass Eva die Studentin einer Gefahr aussetzte
that Eva the$_{ACC}$ student a$_{DAT}$ danger exposed
'that Eva exposed the student to a danger'

The verbs included in the following experiments come from the second group that has IO > DO as its basic word order. They contain prototypical ditransitive verbs and monotransitive verbs to which a dative argument can be added. In Georgala's terminology, these verbs are "high" datives and they contain an applicative head above the lexical VP. This second group is then further subdivided into a thematic applicative (94a) and a raising applicative (95a) that are associated with different thematic roles and different semantic and syntactic behavior. In the thematic applicative the IO is base-generated and receives a theta-role of beneficiary or instrument (examples (94a,b) from Georgala, 2011b, her examples (5) and (12a)).

(94a) [$_{vP}$ SUBJ [$_{v'}$ v [$_{ApplP}$ **IO**$_{BNF/LOC/INSTR...}$ [$_{Appl'}$ Appl [$_{VP}$ V DO]]]]]

(94b) Ich habe dem Kind deinen Rucksack gehalten.
I$_{NOM}$ have the$_{DAT}$ child your$_{ACC}$ backpack held
'I held your backpack for the child.'

The raising applicative is derived through movement of the IO from [Spec, VP] to ApplP and the IO is associated with the thematic role of recipient or possessor (examples (95a,b) are example (6) and (17a) in Georgala, 2011b):

(95a) [$_{vP}$ SUBJ [$_{v'}$ v [$_{ApplP}$ **IO**$_{REC}$ [$_{Appl'}$ Appl [$_{VP}$ **t**$_{IO}$ [$_{V'}$ V DO]]]]]

(95b) Er hat jemandem heimlich einen alten Test gegeben.
he$_{NOM}$ has someone$_{DAT}$ secretly an$_{ACC}$ old$_{ACC}$ test given
'He secretly gave someone an old test.'

A comparison of the schematics in (94a) and (95a) shows that in both cases the IO precedes the DO and the IO ends up above the Appl-node. The raising

---

role of the DP in ApplP changes depending on the type of applicative. Unergative predicates and static verbs can be used to differentiate the two types of applicatives as they can only occur with high applicatives.

applicative contains an additional bracket and a filler-gap dependency that is formed between the raised IO and its original position. Georgala supported her theory by syntactic diagnostics, but experimental evidence for the presence of a trace in raising applicative has not (yet) been presented. An experimental investigation of the distinction between thematic and raising applicatives would depend on the presence of the trace in raising applicatives and the correct perception of the thematic role of the IO. Using the schematic in (95a) and the example in (95b) it is not entirely clear where the trace would be located as the schematic does not assume an adverbial phrase that is present in the example and the applicative (Appl) is not expressed morphologically in German. Assuming that the creation of a filler-gap dependency is inherently costly, no matter how long the distance between the filler and the gap, raising applicatives should elicit longer reading times than thematic applicatives, provided that the reader correctly perceives the IO as either a recipient or a beneficiary. The thematic role of the IO was not an experimental factor in experiment 2b and therefore Georgala's applicative theory cannot be experimentally validated with the data collected. However, there was some variance in the experimental items with regard to the thematic roles, and a post-hoc test of the reading times contrasting verbs assigning a recipient role to the IO with verbs assigning a beneficiary could shed some light on whether there are reading time differences already in the IO > DO order that is assumed to be the canonical order for the verbs investigated.

5.1.2 Background L2 group: Scrambling in the Slavic languages

Speakers from Slavic languages were chosen as the L2 group in this experiment as their native languages show similar properties to German in the area of ditransitive sentences and scrambling (Junghanns & Lenertová, 2007; Siewierska & Uhlírová, 1998). I did not pick one single Slavic language as the majority of Slavic languages are very homogeneous with regard to word order variability and the use of scrambling as a way to mark definiteness in the absence of determiners (e.g. for Russian, King, 1995). The Slavic languages exhibit an amount of free word order that exceeds even the possibilities of word order variation in German, and arguments in non-canonical positions are therefore highly familiar to the L2 group. A review of scrambling in the individual Slavic languages that were present in the L2 group follows below. The small differences between the native languages should not overrule the bigger similarities when it comes to processing, so I do not expect the mixture in native languages to affect the L2 results. As became evident by the preceding review on scrambling in German, a variety of factors are at play that motivate scrambled objects in the midfield. One of these factors, information structure, is central to the scrambling theory in the Eastern European tradition and makes it an interface phenomenon according to Sorace's (2006) Interface Hypothesis. According to this hypothesis, scrambling and the small differences in acceptability that result from it should be hard to acquire for L2 speakers. By using L2 speakers that are familiar with the scrambling and its motivation via information structure already in their L1s, I hypothesized that the task would be more manageable for this L2 group, and the chances of a native-like performance in the L2 might be bigger. Transfer accounts that assume the full or partial transfer of L1 settings would predict that L2 speakers from Slavic languages are able to perceive a difference between the two argument orders and process them differently from each other, as they also mark a difference with regard to information structure and context appropriateness in their native languages. The choice for Slavic speakers in this study was made in order to have a native language that is not closely related to the target language, but nevertheless exhibits scrambling under highly similar circumstances, thereby paving the way for possible positive transfer and presumably making the task easier.

As stated above, Slavic languages have a highly variable word order, but the source of this freedom in ordering is explained differently by various linguistic traditions. Generative grammar accounts assume a derivation of the surface structures via movement that changes the underlying base-generated order ('scrambling'), while the functionalist analysis mainly used in the Soviet, Russian and Prague Schools assumes a close relation between the different word orders and the appropriate contexts in which they appear. These two approaches are not necessarily mutually exclusive, as movement can well be motivated by the demands of the context.

Siewierska & Uhlírová (1998) proposed that the basic unmarked word order for all Slavic languages is SVO as it has the widest contextual applicability and occurs when two arguments are non-distinctive. Statistically this is also the dominant word order with a range of 67 to 88% of all occurrences in corpora of different Slavic languages (see e.g. Siewierska & Uhlírová, 1998 for Polish, Lobanova, 2011 for Russian). With regard to information structure, constituents are usually ordered according to the theme > rheme principle and the rheme > theme order is only used to express a subjective attitude towards the information.

Within the family of Slavic languages there is considerable variation in case systems and the presence of articles. Some languages have no articles and a very elaborate case system of up to seven cases, marked on nouns, pronouns and adjectives (e.g. Serbian and Polish) while others have a limited set of case markers for pronouns and some nouns, but make use of suffixed definite articles (e.g. Bulgarian and Macedonian). These differences have consequences on the word orders used in ditransitive constructions that will be discussed for each subgroup independently below. Unless there are clitics involved, there are no syntactic constraints regarding the order of phrases in main declarative clauses and all possible permutations of subject, direct object and indirect object are grammatical. They are however not freely interchangeable, and neither are they pragmatically or communicatively free. The frequencies of these 24 possible permutations also show a lot of variance in corpora with some orders being extremely infrequent (see Siewierska & Uhlírová, 1998 for more details). In general, it can be said for all Slavic languages that there is no rigid order of patient and recipient, if both constituents are nominals. There are still word order preferences. The case marking languages prefer the order recipient > patient, while the non-case marking languages Bulgarian and Macedonian show a preference for the patient > recipient order. Object clitics that are frequent especially in the South Slavic languages, are always ordered dative clitic > accusative clitic.

The following section familiarizes the reader with the various Slavic languages sorted by subgroup - Eastern, Western and Southern – as well as the experimental evidence that has been gathered to support the theoretical claims stated above. The main sources of empirical evidence are grammaticality judgments that are traditionally used in generative grammar and some corpus studies. Experimental evidence from a processing or production perspective is sparse, but will be reviewed in this section as well.

East Slavic (Russian, Ukrainian, Belarusian)

The East Slavic languages Russian, Belarusian and Ukrainian have a rich morphological case system. Neither of the three languages has articles, and the order of objects in ditransitive sentences is among other things used to express definiteness and focus (Junghanns & Lenertová, 2007). The DAT > ACC order involves focus ambiguity as it allows an indefinite or a definite interpretation of the dative object (96a). The ACC > DAT order is used in more limited contexts with narrow focus and allows only one interpretation: a definite accusative object with an indefinite dative object (96b). Due to the wider range of possible contexts in which it can occur, the DAT > ACC order is usually considered the more neutral one from which the ACC > DAT order is derived via movement to fit the needs of the informational structure of the sentence, but see Bailyn (2010) for an opposite view. The examples below are taken from Mykhaylyk, Rodina, & Anderssen (2013):

(96a) Petja dal devočke knigu. Russian
Petryk dav divčynci knyhu. Ukrainian
Peter$_{\text{NOM}}$ give$_{\text{PRET}}$ girl$_{\text{DAT}}$ book$_{\text{ACC}}$
'Peter gave a/the girl a book.'

(96b) Petja dal knigu devočke. Russian
Petryk dav knyhu divčynci. Ukrainian
Peter$_{\text{NOM}}$ give$_{\text{PRET}}$ book$_{\text{ACC}}$ girl$_{\text{DAT}}$
'Peter gave the book to a girl.'

In their elicited production study on word order alternations in Russian and Ukrainian, Mykhaylyk et al. found that adults in both languages respected the givenness of recipient or theme when producing their sentences. In contexts in which the recipient was given, adults produced the DAT > ACC order, and in theme-given contexts the ACC > DAT order. The children that were also tested in this experiment (between the ages of 3 and 6) produced the DAT > ACC order more frequently irrespective of the information structure. The authors explained the children's results with a general preference for the less complex, movement-free DAT > ACC order in children and not by a lack of sensitivity to the information structure. In an earlier study on the production of word orders in adult Russian, Kallestinova (2007) had found the same strong preference to place focused arguments at the end of a sentence. In the condition with a focused DO, the basic word order S V IO DO counted for 71.2% of all productions while the other basic order S V DO IO was not produced at all. The reverse effect (79.9% against 0%) was found for the S V DO IO order in the IO focus condition. However, Kallestinova had not found a production preference for DAT > ACC or ACC > DAT in sentences without focus as both orders occurred at roughly similar rates – 48.9% for the DAT > ACC order and 46.4% for the ACC > DAT order. In contexts that clearly focus one argument, Russian native speakers adhere to the unfocused > focused ordering principle. In contexts without a focused argument, the suggestion of the functionalist analysis that the DAT > ACC order should be preferred, could not be supported by production experiments.

Slioussar (2011) investigated ditransitive sentences that either had their canonical argument order S V IO DO (97a), contained a fronted IO (97b) or DO (97c), or contained two fronted objects (97d + e). In order to assess the effects of context on the processing of non-canonical orders, the sentences were presented either with an appropriate or an inappropriate context, using a match/mismatch design. An appropriate context always mentioned the first two constituents of its matching sentence, thereby rendering the last constituent new information. The context for sentence (97b) below was *On March 8, Danya Kashin bought a plush elephant and a box of candies.*

(97a) Danja Kašin podaril Maše Smolinoj pljuševo slona.
Danya$_{\text{NOM}}$ Kashin$_{\text{NOM}}$ gave Masha$_{\text{DAT}}$ Smolina$_{\text{DAT}}$ plush$_{\text{ACC}}$ elephant$_{\text{ACC}}$
'Danya Kashin gave Masha Smolina a/the plush elephant.'

(97b) Pljuševo slona Danja Kašin podaril Maše Smolinoj.
plush$_{\text{ACC}}$ elephant$_{\text{ACC}}$ Danya$_{\text{NOM}}$ Kashin$_{\text{NOM}}$ gave Masha$_{\text{DAT}}$ Smolina $_{\text{DAT}}$

(97c) Pljuševo slona Maše Smolinoj podaril Danja Kašin
plush$_{\text{ACC}}$ elephant$_{\text{ACC}}$ Masha$_{\text{DAT}}$ Smolina$_{\text{DAT}}$ gave Danya$_{\text{NOM}}$ Kashin$_{\text{NOM}}$

(97d) Maše Smolinoj Danja Kašin podaril pljuševo slona.
Masha$_{\text{DAT}}$ Smolina$_{\text{DAT}}$ Danya$_{\text{NOM}}$ Kashin$_{\text{NOM}}$ gave plush$_{\text{ACC}}$ elephant$_{\text{ACC}}$

(97e) Maše Smolinoj pljuševo slona podaril Danja Kašin.
Masha$_{\text{DAT}}$ Smolina$_{\text{DAT}}$ plush$_{\text{ACC}}$ elephant$_{\text{ACC}}$ gave Danya$_{\text{NOM}}$ Kashin$_{\text{NOM}}$

The SPR data showed a general effect of context mismatch, as sentence-initial new information that had not been introduced by the context lead to increased reading times compared to control sentences that contained sentence-initial given information. The immediate occurrence of this slowdown was interpreted by

Slioussar as evidence for immediate access and integration of context information by the native processor. Whether this slowdown continued beyond the first element to also affect the following given constituents depended on the canonicity of the following constituent. Sentences like (97b) and (97d) that continued with a subject in its canonical preverbal position showed a significant difference between faster reading times in the context match condition compared to a slowdown in the context mismatch condition. The same was true for the IO in sentence (97a) that caused slower reading times for mismatched context than for matched context. When the following constituent was an object in a non-canonical position, the IO in (97c) or the DO in (97e), there small slowdown in reading times for the context mismatch condition compared to the context match condition was not significant and neither was the numerical difference at the following verb. Slioussar interpreted this finding to reflect the psycholinguistic reality of scrambling, as scrambled and unscrambled constituents behaved differently. Overall, an appropriate context facilitated reading time speed for sentences with a single instance of scrambling (97b+d) as well as for sentences with multiple scrambling (97c+e) and a slowdown in reading times was caused by the unexpected appearance of new information in a sentence-initial instead of a sentence-final position. With the design used by Slioussar, the context mismatch condition for non-canonical word orders always used a context that presupposed the canonical order, but never a context that could have motivated a different kind of non-canonical order which could explain the absence of an effect for the non-canonical second constituents in (97c) and (97e). The canonical sentences occurred in two mismatch conditions that motivated two types of non-canonical word orders and also yielded different results for the canonical IO in (97a). An extension of the design to include a second mismatch condition for the non-canonical conditions might have been appropriated to reveal the full extent of context on the processing of scrambled objects.

With regard to focus, Neeleman & Titov (2009) propose two different options for focus positions: new information foci are grammatical only in a clause-final position (98a), while contrastive foci appear in a fronted position (98b). In both cases, the focused information receives the main stress of the sentence.

(98a) Kto dala Kate knigu?
Who$_{NOM}$ give$_{PRET}$ Katja$_{DAT}$ book$_{ACC}$
'Who gave Katja a book?'

Kate knigu dala Anja
Katja$_{DAT}$ book$_{ACC}$ give$_{PRET}$ Anja$_{NOM}$
'Anja gave Katja a book.'

(98b) Anja dala Kate knigu?
Anja$_{NOM}$ give$_{PRET}$ Katja$_{DAT}$ book$_{ACC}$
'Did Anja give Katja a book?

Net, Olga dala Kate knigu.
No, Olga$_{NOM}$ give$_{PRET}$ Katja$_{DAT}$ book$_{ACC}$
'No, Olga gave Katja a book.'

Lobanova's (2011) corpus study looked at the role that animacy and definiteness play in the assignment of subjecthood in Russian. Previous studies on languages with comparatively fixed word orders such as English and Swedish (Zeevat & Jäger, 2002) or German (Heylen, 2005) had found that the close association of animacy and definiteness with grammatical functions is strong enough to have predictive power. NPs that were animate and/or definite were highly likely to be subjects, while inanimate and/or indefinite NPs tended to be objects. Lobanova analyzed a set of 600 SVO and OVS sentences in order to find similar associations of animacy and definiteness with grammatical function. She found that unlike in the previously studied languages, animacy and definiteness in Russian interact with information structure rather than with grammatical function. While the expected strong association between animacy and grammatical function was found in the SVO sentences, it was weakened in the OVS sentences as the number of animate objects increased. Overall, non-typical elements in marked positions were avoided less than in the previously investigated languages with fixed word orders. With regard to referential form, proper names were the only reliable predictor of subject function independent of word order. Pronominalization was a reliable predictor only in canonical SVO sentences, as the majority of pronouns in OVS sentences referred to the object, again occupying a marked position in the sentence. Unlike in previously investigated languages, animacy and definiteness in non-canonical word orders cannot be used to predict the grammatical function of a constituent in Russian. In terms of linearization preferences this means that while the Animacy and Definiteness constraint were found to be closely tied to the

Dative constraint in German, they might be more closely associated with the Givenness constraint in Russian.

West Slavic (Polish, Czech, Slovak)

Similarly to the East Slavic languages, the West Slavic languages Polish, Czech and Slovak have a rich morphological system without articles. In the following paragraph theoretical approaches to word order variation in Polish and Czech will be discussed.

Siewierska (1993) used corpus data on Polish word order variation to compare the predictive strength of the performance-based principle EIC (see above) to the traditional pragmatics-based approaches. The corpus data showed a 72.5% frequency of the SVO order in sentences in which subject and object were overtly expressed. While the EIC predicts this dominance of the SVO order, as it is the word order associated with the lowest processing effort, it is unable to predict the occurrence of any of the other five word orders (SOV, VSO, VOS, OVS, OSV) that are grammatical and can be found in the corpus, but should always lose to the SVO order due to higher processing effort. Any changes to the SVO order can therefore not be motivated solely based on the EIC principle. According to Siewierska, the Topic > Comment principle can describe the state of Polish word order variation more accurately. The predictable information that is part of the topic changes with the word order type. In SVO and OVS orders it usually consist only of the first constituent, but can be extended to the verb. In SOV and OSV orders, the topic/comment division always occurs after the first constituent and in VSO and VOS orders the topic consists of the first two constituents. Siewierska also states that the topic/comment structure of a sentence in context is reliably identified by native speakers and is part of acquired knowledge due to a highly prescriptive tradition of teaching the topic > comment order in written Polish. Unfortunately, Siewierska does not make any claims about the ordering of objects in double object structures that are investigated in the following experiment.

Biskup (2006) analyzes the properties of scrambling in Czech in the domains of semantics and information-structure within a Minimalist Framework. Scrambled subjects and objects necessarily have to receive a specific interpretation and are considered backgrounded information. However, subjects and objects in situ can receive either an existential or a specific reading and are interpreted as informationally focused. In Czech, dative and accusative objects are equally able to scramble (99a and b) and can also scramble in the same sentence (99c). In this case both orders of scrambled objects are equally grammatical. In example (99c) below, the order DAT > ACC is exemplified, the order ACC > DAT (…dopisy detem…) is possible as well.

(99a) Pavel detem odpoledne poslal dopisy.
Pavel$_{\text{NOM}}$ children$_{\text{DAT}}$ in the afternoon send$_{\text{PRET}}$ letters$_{\text{ACC}}$
'Pavel sent the children letters in the afternoon.'

(99b) Pavel dopisy odpoledne poslal detem.
Pavel$_{\text{NOM}}$ letters$_{\text{ACC}}$ in the afternoon send$_{\text{PRET}}$ children$_{\text{DAT}}$
'Pavel sent the letters to children in the afternoon.'

(99c) Pavel detem dopisy odpoledne poslal.
Pavel$_{\text{NOM}}$ children$_{\text{DAT}}$ letters$_{\text{ACC}}$ in the afternoon send$_{\text{PRET}}$
'Pavel sent the children the letters in the afternoon.'

Biskup unfortunately does not give any frequency counts for the two scrambled orders nor does he suggest a basic word order for non-scrambled objects.

Dvořák (2010) agrees with Biskup's analysis in so far that she considers both argument orders (DAT > ACC or ACC > DAT) as grammatical in Czech, but she argues for two distinct classes of ditransitive verbs similarly to Georgala's analysis of German above. The verbs have two different base-generated orders in neutral contexts: either DAT > ACC or ACC > DAT. Table 5.4 below gives examples of DAT > ACC and ACC > DAT verbs. As both constituent orders are grammatical in Czech, ACC > DAT is the unmarked order for ACC > DAT verbs, but the marked order for DAT > ACC verbs. Dvořák also analyzes the dative differently for these two verb types: in DAT > ACC verbs, the dative is an applicative verbal head with a thematic role of benefactive or recipient, in ACC > DAT verbs the dative is a PP complement with a phonologically null preposition and a thematic role of path.

DAT > ACC		ACC > DAT
with recipient meaning	with benefactive/ malefactive meaning	
dát ‘give’	věřit ‘believe’	vystavit ‘expose’
darovat ‘give as a gift’	vařit ‘cook’	zasvětit ‘devote’
postlat ‘send’	ukrást ‘steal’	svěřit ‘entrust’
vrátit ‘return’	ztratit ‘loose’	podřídit ‘subordinate, accommodate’
zakázat ‘forbid’	popřit ‘deny’	podrobit ‘put sb. through sth.’

Table 5.4Classification of Czech ditransitive verbs based on Dvořák (2010)

The two verb groups differ with respect to their behavior under topicalization, passivization and the obligatoriness of the dative. There is also a difference between the two verb groups with regard to productivity and what type of nouns can occur as datives. ACC > DAT verbs are a small closed class and can take animate and inanimate nouns as datives. DAT > ACC verbs are highly productive, but they are limited to animate datives. Biskup’s and Dvořák’s approaches can be reconciled as Biskup’s analysis seems to be mainly based on what Dvořák would classify as DAT > ACC verbs.

South Slavic (Slovenian, Croatian/Serbian, Bulgarian)

The South Slavic languages fall into two groups: Slovenian, Croatian, Bosnian, Serbian and Montenegrin (the latter four being previously subsumed under the name ‘Serbo-Croatian’) form one group that shares the presence of a rich case marking system and absence of articles. Bulgarian and Macedonian make up another group which features restricted case marking and the presence of a suffixed definite article. Interestingly enough, despite these differences, similarities in the theoretical approaches towards argument order in ditransitive sentences can be found. Slavkov (2008) for Bulgarian, Marvin & Stegovec (2012) for Slovenian and Gračanin-Yuksek (2006) for Croatian all analyze the two argument orders in terms of the dative alternation found in English. However, only Bulgarian and Macedonian possess a dative that involves a preposition like the English *to* dative used in the Prepositional Dative Construction (PDC). In the other languages the dative alternation is an alternation of the order of the two object arguments without additional syntactic material. The Double Object Construction (DOC) is identified using standard syntactic diagnostics such as the availability of a causative reading, nominalizations, binding properties and frozen or free quantifier scope.

In Bulgarian, there are three ways to express an object: either through a full NP (100a), a case-marked clitic (100b), or an NP and an additional clitic also known as ‘clitic doubling’ (100c). The indirect object is printed in bold in the following examples (adapted from Slavkov 2008:141)

(100a) Ivan izprati pismo **na Marija**.
Ivan send$_{\text{PRET}}$ letter to Marija.
‘Ivan sent a letter to Marija.’

(100b) Ivan **i** izprati pismo.
Ivan CL$_{\text{DAT}}$ send$_{\text{PRET}}$ letter
‘Ivan sent her a letter.’

(100c) Ivan **i** izprati pismo **na Marija**.
Ivan CL$_{\text{DAT}}$ send$_{\text{PRET}}$ letter to Marija.
‘Ivan sent Marija a letter.’

The preferred order of objects in Bulgarian depends on how these objects are realized. The basic word order for two full NPs as objects is DO > IO (100a), but in a clitic cluster the order is reversed to DAT > ACC (Junghanns & Lenertová, 2007). These orders are not rigid, though, and allow a flexible word order if both objects are full NPs as in (100a) and (100c). Slavkov (2008) suggests the existence of a DOC in Bulgarian that is created in the presence of clitic doubling. The introduction of the clitic into sentence (100c) changes its behavior with regard to possessive binding, weak crossover and frozen scope compared to (100a) which Slavkov takes as evidence for a structural difference between the two sentences: (100a) is to be analyzed as a typical PDC in which the Theme c-commands the Goal that is expressed by a directional PP, while (100c) is a DOC that contains an applicative phrase with the Goal as its specifier, the dative clitic in the head and the Theme as a complement. All non-clitic dative objects have to be accompanied by a preposition-like *na*. In the PDC *na* is a meaningful PP corresponding to the *to*-phrase in English, while Slavkov interprets the *na* as a dative case marker in the DOC.

For Slovenian, Marvin and Stegovec (2012) state that the two object orders in ditransitive sentences are realizations of two different base-generated, deep structures with different properties. They also argue that the displacement of arguments based on information structure is distinct from word order alternations in ditransitive sentences. Ditransitive sentences with specific interpretations retain these readings also after displacements, while the interpretations of reversed arguments in neutral environments do not retain their original reading. In information-structure neutral contexts, the DAT > ACC order allows a high (101a) and low applicative (101b) or a causative reading, while the ACC > DAT order only allows a low applicative (101c) reading and is therefore more similar to the English prepositional dative. Note from example (101c) below that while Marvin & Stegovec interpret the ACC > DAT order like a PDC, there is no overt preposition. The dative DP is analyzed as a complement of a phonologically null prepositional head P.

(101a) Bine pošilja Zoji pismo.
Bine$_{\text{NOM}}$ send$_{\text{PRES}}$ Zoja$_{\text{DAT}}$ letter$_{\text{ACC}}$.
high applicative: 'Bine is sending a letter for Zoja.' (because Zoja cannot do it)

(101b) low applicative: 'Bine is sending Zoja a letter.'

(101c) Bine pošilja pismo Zoji.
Bine$_{\text{NOM}}$ sending letter$_{\text{ACC}}$ Zoja$_{\text{DAT}}$.
low applicative: 'Bine is sending a letter to Zoja.'

Gračanin-Yuksek's (2006) analysis of Croatian comes to a similar conclusion as Marvin and Stegovec' (2012) analysis of Slovenian. The order dative-verb-accusative (DVA) is classified as a double object construction and the VAD order as the prepositional dative. The VDA order is interpreted as ambiguous between the two structures and sharing properties of both in Croatian, while Marvin and Stegovec (2012) interpret this order as a clear double object construction like the VAD order in Slovenian. Gračanin-Yuksek also suggests different structures for the three word orders, but argues against an applicative head that was suggested in the analysis of Marvin and Stegovec (2012) and Slavkov (2008).

In summary, even though there are differences between the Slavic languages and experimental evidence is spare for some of them, there is a general consensus that changes in word order also reflect a change in sentence content and that native speakers are sensitive to this change. Table 5.5 provides a summary of the ordering preferences in German and the Slavic languages for the types of verbs investigated in Experiments 2a and 2b.

	Canonical order	**Non-canonical order**
German	DAT > ACC (full NPs) ACC > DAT (pronouns)	ACC > DAT (full NPs) DAT > ACC (pronouns)
Russian Ukrainian Polish Czech Slovenian Croatian	DAT > ACC	ACC > DAT
Bulgarian	ACC > DAT (full NPs) DAT > ACC (pronouns)	DAT > ACC (full NPs) ACC > DAT (pronouns)

Table 5.5 Cross-linguistic overview of ordering preferences in ditransitive sentences

With the exception of Bulgarian that has the exactly opposite ordering preferences of German regarding full NPs and pronouns, the Slavic languages show a uniform preference for the DAT > ACC order as the canonical object order in ditransitive sentences, just like German. A non-canonical ordering is always possible in all of the Slavic languages, also Bulgarian.

Important theoretical questions remain regarding the motivation of the change from canonical to non-canonical word order. One possibility is the traditional approach that interprets changes in word order as a means to highlight the expression of topicality, definiteness and focus. Other theories focus on differences on a lower level such as the expression of different theta roles of the dative argument or different preferences for individual verbs. However, these questions are outside the scope of this dissertation. As neither the theta role of the dative nor the order preference of the verbs was manipulated in Experiments 2a and 2b, I adopt the traditional approach to word order changes for this thesis.
In their native languages, my L2 German speakers seem to be highly sensitive to differences between related argument orders, and Study 2 investigated whether they show this sensitivity also in their German L2 with the following research questions.

Research questions:

Q2.1 Do L1 and L2 speakers of German perceive one object order as more acceptable than the other in no context situations? Is there evidence for a gradience in acceptability in both groups?

Q2.2 Are canonical word orders processed more easily than non-canonical orders? Is a processing advantage visible in native and non-native speakers alike?

Q2.3 Is the L2 result possibly influenced by the variance in Age of Acquisition, the time spent learning German and the length of stay in Germany that varied considerably between the participants?

5.2 Experiment 2a: Acceptability rating

This task was administered to check whether the participants had an offline preference for one order of objects over the other in sentences without context and whether this preference would be the same in both groups.

Participants

Altogether 74 participants took part in this experiment and received either course credit or 8€ for their participation. Participants were mainly recruited and tested in the Berlin-Brandenburg area.

The L1 control group consisted of 33 native speakers of German who were mainly students at the University of Potsdam or the Humboldt-University Berlin. The group (male n=7) had an average age of 23.88 years (SD: 3.9, range: 18-39). Three participants were left-handed. All participants had normal or corrected-to-normal vision. All participants were monolingual native speakers of German and had learned English as a foreign language later in life. No language disorders were reported.

The original L2 group consisted of 41 German L2 speakers with a variety of Slavic languages as their L1. They were mostly students of the University of Potsdam, either enrolled in programs at the university or taking part in exchange programs. Two participants were excluded as they reported either having grown up bilingually with German from birth or had a very low AoA for German (at 3 years) and had been born in Germany, suggesting that the Slavic L1 (Serbian) was probably a heritage language. Four other Russian-speaking participants were Russian-Germans who had come to Germany at an age between 5 and 8 and started learning German upon arrival. Their data will be flagged as they could also be possible heritage speakers and if differences from the rest of the group emerge, they will be reported. The data of several other participants was excluded from subparts of the experiment as they did not pass certain cutoff criteria. These additional exclusions are reported in the results sections of the corresponding subpart.

The remaining group of 39 participants (male n=8) had an average age 25.7 (SD: 3.9, range: 20-34). All but four participants were right-handed. All had normal or corrected-to-normal vision with two participants reporting squinting. No language impairment or neurological disorders were reported. The Ukrainian participants were bilingual speakers and reported having grown up with both Ukrainian and Russian from birth. One Czech speaker indicated that she had grown up bilingually with Slovak as a second language.

All participants completed the Goethe placement test (Goethe Institute, 2010).[11] The L2 group scored 23.82/30 (range: 12-30, SD: 4.03) corresponding to an average B2 level in the Common European Reference Framework for Languages (CERF), range was B1 to C2. All participants in the L2 group can be classified as at least intermediate with the majority being advanced speakers.

Average AoA for German in the L2 group was 14.03 years (SD: 6.26, range: 5-32) and average length of learning German was 11 years (SD: 5.35, range: 1.5–23). Eleven participants had started to learn German in or before elementary school (AoA 10 years or younger), but the majority of participants had started to learn German after the age of 10 and in a school setting. For 13 participants German was the first foreign language and for another 21 German was the second foreign language, usually after English. Average age of arrival in Germany was 20.26 years (SD: 6.48, range: 5-32). Length of stay varied considerably between the participants as some were permanent residents of Germany, others were enrolled in study programs spanning several years or were exchange students spending one or two semesters at the University of Potsdam. The native languages of the L2 group were highly mixed with about half the group (n=21)

[11] This test is a multiple-choice test with 30 items used by the Goethe institute, a worldwide institution supported by the German government that provides language classes and promotes German culture, to broadly assess the proficiency level of a speaker for placement in its classes. The 30 items are ordered by increasing difficulty and the test-taker has four options to choose from, and additional fifth one being “I don’t know.” The test is focused on reading comprehension, the application of specific grammatical rules and lexical knowledge. Speaking or listening ability are not tested. I used a paper-based version of the test that was usually completed within 10-15 minutes by the L2 participants. The test does not query the structures investigated in experiment 2 and 4. It can at best give an idea of the language abilities of the test-taker, but should not be equated to an exhaustive proficiency test.

having Russian as (one of) their native language(s). Of the remaining participants, 12 came from the Western group of Slavic languages (Polish and Czech) and 6 from the Southern group of Slavic languages (Bulgarian, Macedonian, Serbian and Slovenian). Apart from the bilingual participants mentioned above, seven participants also had additional experience with another Slavic language as a foreign language (three had learned Russian, three had learned Ukrainian and one had learned Croatian). Table 5.6 provides an overview of the biographical data of the L2 group split according to L1 background. The group was mostly homogenous regarding their AoAs and years studying German. The Macedonian and Serbian speakers differ from the rest of the group with a comparatively high AoA and short learning time. Due to the small sample size for most of the L1 backgrounds, no statistical analysis was performed to check whether any group differed statistically from the others. With sample sizes this small, individual values affect the averages too much and a normal distribution that is assumed in most statistical tests is not given.

L1	**N**	**Age**	**Goethe score**	**AoA German**	**Years of German**
Bulgarian	2	28.5	25.5	14	11.5
Czech	6	24.7	21.2	13.2	10.3
Macedonian	1	29	19	28	1.5
Polish	6	24.7	24.2	14.8	11.7
Russian	19	25.9	24.7	12.6	12.8
Serbian	1	34	24	32	2
Slovenian	2	23	26	9.5	13.5
Ukrainian	2	24.5	21	16	6

Table 5.6 Biographical data of the L2 group split according to native language

Materials

The acceptability rating task employed a simple design with object order as its only factor with two levels: DAT > ACC (canonical), ACC > DAT (non-canonical). There were 8 experimental sentences that were presented without a context. Each sentence featured an animate subject, an animate dative object and an accusative object that was mostly inanimate. Both objects were definite as this is the most frequent pattern for rearranged orders (Kurz, 2000) and should have the least impact on the acceptability of the accusative-first orders. The objects were matched for frequency ($t(11.245)=-0.21$, $p=0.84$)) and although the accusative objects were slightly shorter on average than the dative objects (7.25

vs. 8.625 respectively), this difference was not significant ($t(8.51)=1.64$, $p=0.14$). All sentences were written in present perfect. Participants saw half the sentences in a dative-first order, the other half in an accusative-first order. These sentences were all grammatically correct and supposed to be rated as acceptable.

DAT > ACC
(102a) Die Enkelin hat dem Großvater den Kuchen mitgebracht.
the$_{NOM}$ granddaughter has the$_{DAT}$ grandfather the$_{ACC}$ cake brought.
'The granddaughter has brought the grandfather the cake.'

ACC > DAT
(102b) Die Enkelin hat den Kuchen dem Großvater mitgebracht.
the$_{NOM}$ granddaughter has the$_{ACC}$ cake the$_{DAT}$ grandfather brought.
'The granddaughter has brought the cake to the grandfather.'

Mixed with these eight sentences were 24 sentences modeled after items from Study 4 (see Section 8.2) and 18 true fillers. Of these true fillers four were grammatically correct transitive sentences in present perfect tense, six were grammatically incorrect present tense sentences that featured a prefixed verb and an additional particle that usually fit with the verb stem. The remaining eight fillers were again transitive sentences in present perfect tense. These sentences used homographic verbs that could either be prefixed verbs or particle verbs (*'wiederholen* 'to take back' vs. *wieder'holen* 'to repeat') and the decision regarding acceptability had to be based on the semantics of the verb and the object. All eight fillers of this type were grammatical from a syntactic point of view, but were unacceptable from a semantic point of view as the object did not fit with the verb.

Altogether participants rated 50 sentences. Thirty of these sentences were written in the simple present, and 20 in the present perfect. Twenty-four sentences were grammatically correct and 26 sentences were grammatically incorrect.

Procedure

One experimental session included the acceptability rating task with items from Experiments 2a and 4a, and the SPR task including items from Experiments 2b and 4b. At the beginning of the experimental session, each participant filled out a questionnaire with biographical information and was handed an information sheet detailing the procedure of the following experiment, the handling of the data, and the respect of personal information. Participants then gave their informed consent to participate in the experiment that could be withdrawn at any point. The participants then completed the SPR task followed by a short vocabulary list for the L2 group. Participants were asked to indicate any unfamiliar words from this list. More detailed specifications for the procedure of the SPR task and the vocabulary list are reported in Section 5.3. The participants were then given the acceptability rating task. It was an untimed pen-and-paper questionnaire. The instructions told the participants that they should judge the acceptability of the sentences to follow. Acceptability was defined as sounding correct and natural. Participants were asked to give their rating on a Likert Scale from 1 to 5, with 1 being acceptable and 5 being unacceptable. Two examples were given to illustrate the end points of the scale. The acceptable example was the sentence *Julia trinkt ihr Glas Milch aus.* 'Julia drinks her glass of milk up.' in which the particle *aus* has been correctly split from the main verb in a V2 context. The unacceptable example **Annika deckt ein neues Land ent.* 'Annika discovers a new country.' contains the prefix *ent-* that has been incorrectly split from the verb stem. These examples used the structure of the critical items for Study 4 (Section 8.2). As the object order tested in this task is by itself always grammatical, I did not want to bias participants by presenting one order as the acceptable example and a completely unrelated phenomenon as the unacceptable example. Participants were also told that they were free to indicate the source of the unacceptability in the sentence and there was enough space in the questionnaire to mark the error or write down corrections. About half of all participants did make corrections to the sentences. The sentences were written in Verdana font size 8. The numbered scale was provided and its ends labeled. Participants were instructed to circle the number that corresponded to their rating. Figure 5.3 below shows the instructions and the first four sentences of the acceptability rating. Item 1 and 3 are fillers, item 2 is an item from Experiment 4b, and item 4 represents an item from Experiment 2b.

In diesem Fragebogen sollen Sie bewerten wie akzeptabel die folgenden Sätze sind. Akzeptabel bedeutet hier, dass sie korrekt und natürlich klingen. Es kann die gesamte Skala ausgenutzt werden. Beispiel: Julia trinkt ihr Glas Milch aus. -> akzeptabel = 1 Annika deckt ein neues Land ent. -> inakzeptabel =5	**Skala** **akzeptabel**				**inakzeptabel**
1. Der Fährmann hat die Fähre übersetzt.	1	2	3	4	5
2. Das Kind isst seinen Brei auf.	1	2	3	4	5
3. Die Journalistin hat den Begriff umgeschrieben.	1	2	3	4	5
4. Der Sänger hat dem Publikum den Song vorgesungen.	1	2	3	4	5

Figure 5.3 First four items of the acceptability rating task used in Experiments 2b and 4b

The critical sentences were spread across two lists using a Latin Square design and the order was pseudorandomized. Two additional lists presented the items in a reversed order. The completion of this questionnaire took 10 to 15 minutes. The completion of the German Goethe placement test was the last task for both groups. It was also untimed and participants were not allowed to ask any questions concerning the content of the test. The L2 group took up to 20 minutes to complete this test. The entire experimental session took between 20 and 30 minutes for the L1 group and between 45 and 60 minutes for the L2 group.

Predictions

The predictions for the acceptability rating task are based on the acceptability of object orders in sentences without context and on the underlying syntactic structures of the two object orders. An overview of idealized ratings based on the predictions can be found in Table 5.7 below.

First of all, both orders should be rated as acceptable as neither order is ungrammatical without a context, although one might be more unusual or infrequent than the other. Previous research eliciting acceptability ratings from native German speakers showed only small differences between the two orders, with both orders considered acceptable overall (Keller, 2000). On a rating scale with 5 as the most extreme ‘unacceptable’ rating, I do not expect average ratings for either condition to go beyond 3. Any rating averages beyond this point will be inspected separately to check the source of this unexpected rating (e.g. flipped scale).

A – The DAT > ACC order should be rated more acceptable than the reversed ACC > DAT order, because…

- it is the more frequent word order for ditransitive sentences involving two nominal NPs (Kempen Harbusch, 2005),
- it is the unmarked order that can appear in context-free situations with a wide focus structure as it is the case in this task,
- it also follows a possible animacy hierarchy, as dative objects were always animate and accusative objects mainly inanimate,
- it does not violate the Dative constraint.

The exact magnitude of the difference might depend on whether the participants experience strong processing problems in the ACC > DAT condition. Previous studies (e.g. Bader & Meng, 1999; Fanselow & Frisch, 2006; Keller, 2000) had found that strong processing problems negatively affect the perceived acceptability of a sentence. A comparatively small rating difference might indicate also little additional processing effort associated with the non-canonical order.

L2 predictions

B – If the L2 group transfers the ordering preferences from their different L1s, I also expect the L2 group as a whole to rate the DAT > ACC order as more acceptable as it is the preferred order in context-free sentence with wide focus. The three speakers with Bulgarian and Macedonian might show the opposite tendency as their languages have ACC > DAT as the canonical word order for full NPs. Given the small number of Bulgarian and Macedonian participants, their results should not impact the overall group result.

C – If the L2 group is more sensitive to differences in frequency of occurrence than the L1 group, the difference between the two ratings should be bigger than for the L1 group and disfavor the ACC > DAT order more strongly due to its much lower frequency. The L2 group might be less familiar with the non-canonical order and therefore possibly perceive it as ungrammatical or at least less acceptable than the more frequent canonical order. If the L2 group is stricter in their application of the animacy hierarchy or the Dative constraint that both favor the canonical order, the same disadvantage for the non-canonical ACC > DAT would also emerge,

D – If the L2 group is insensitive to frequency differences, focus structure and other possible linearization principles or at least does not link them to acceptability, both orders should be equally acceptable.

	Rating DAT > ACC	**Rating ACC > DAT**
A – expected native behavior based on previous research	1	2
B – transfer of L1 preferences[12]	1	2
C – more frequent order more acceptable	1	4
D – no application of linearization principles, nor frequency	1	1

Table 5.7 Idealized rating averages for Experiment 2a

Results

The analysis of the full data set with ratings from all 72 participants (33 L1 and 39 L2) contained 575 data points. One data point was missing as one participant did not provide an answer for one item. Table 5.8 below shows that both orders were on average rated as acceptable by both groups with average ratings well below 2. The ratings of the ungrammatical items from experiment 4a with ratings of 4.3 and 4.4 showed that the participants correctly judged ungrammatical items as not acceptable and used the full range of the rating scale. The rating for the filler sentences also showed that participants had used the full scale to assess the sentences. Grammatically correct filler sentences had received ratings of 1.14 by the L1 group and 1.58 by the L2 group, while the grammatically incorrect sentences were rated with 4.37 and 4.08 respectively. The semantically awkward fillers received ratings that were more in the middle of the acceptability scale at 3.54 and 3.15. The average ratings for the ACC > DAT sentences were highly

[12] The three Bulgarian and Macedonian participants would show the opposite pattern with a better acceptability of the ACC > DAT order.

similar in the L1 and the L2 group. They were numerically higher than their respective DAT > ACC ratings and also showed more variance in the ratings in the L1 group. The dative-first ratings exhibited a floor effect in the L1 group with a highly acceptable rating and only very few ratings deviating from 1. The L2 group showed a lot more variance in the ratings for the dative-first condition than the L1 group and also had a higher average rating of 1.61 for this condition. The difference between the dative-first order and the accusative-first order was five times bigger in the L1 group (0.6) than in the L2 group (0.12).

	DAT > ACC	**ACC > DAT**
L1 group	1.17 (0.35)	1.77 (0.67)
L2 group	1.61 (0.8)	1.73 (0.83)

Table 5.8 Average acceptability ratings across conditions, SD in brackets, full dataset (Experiment 2a)

A between-groups ANOVA showed a significant main effect of Order ($F_1(1,70)=12.6$, $p<0.001$, $F_2(1,7)=9.72$, $p=0.017$) and a Group x Order interaction that was significant only in the by-subjects analysis and marginally significant in the by-items analysis ($F_1(1,70)=6.2$, $p=0.015$, $F_2(1,7)=4.84$, $p=0.064$). The data of the two groups was then analyzed separately and paired t-tests showed a significant effect of Order in the L1 group ($t_1(32)= -5.8$, $p<0.001$, $t_2(7)= -3.6$, $p=0.009$), but not in the L2 group ($t_1(38)= -0.78$, $p=0.44$, $t_2(7)= -0.75$, $p=0.5$). The numerical advantage for the dative-first order seen in table 21 above is only a robust preference in the L1 group. When considering the entire dataset, the L2 group does not have a robust preference for one order over the other.

Post-hoc analyses

Unlike in the usual case in which acceptability data is collected only through ratings, participants in this task had the possibility to comment on and correct the given sentences which allows further insight into the data. The number of corrections was similar in both groups: the L1 group overall provided 49 corrections and the L2 group provided 47 corrections. The types of corrections were different though and can be seen in Table 5.9 below.

	L1 group (N=33)	L2 group (N=39)
Change ACC > DAT to DAT > ACC	40 (81.6%)	11 (23.4%)
Change DAT > ACC to ACC > DAT	4 (8.16%)	7 (14.89%)
Unrelated to manipulation	5 (10.2%)	29 (61.7%)

Table 5.9 Distribution of corrections of experimental items in L1 and L2

Corrections that were unrelated to the experimental manipulation were, for instance, the deletion of an article to produce a mass noun or changing the verb or a noun, among other possibilities. 19 of the unrelated corrections in the L2 group concerned changes to the verb. Three verbs seemed especially problematic for the L2 group, namely *vortragen* 'to recite', *ausreden* 'to talk someone out of something' and *ausgeben* 'to buy someone a drink'. The corrections provided by the participants show a difference between the two groups: while the L1 group marks the preference for the dative-first order while still giving good ratings to the accusative-first order, the L2 group focuses more on parts of the sentence that were not part of the manipulation.

As the data is available, I decided to remove all those data points in which the correction and therefore likely also the rating given to the sentence was unrelated to the manipulation. This resulted in the removal of five data points from the L1 data (1.9%) and 29 data points from the L2 data (9.3%), reducing the dataset to 541 data points.

The between-groups ANOVA then only showed a significant main effect of Order ($F_1(1,70)=21.9$, $p<0.001$, $F_2(1,7)=15.3$, $p=0.006$). The Group x Condition interaction was only marginally significant ($F_1(1,70)=3.13$, $p=0.08$, $F_2(1,7)=4.31$, $p=0.076$). As more data was removed in the L2 group, bigger changes are visible in this group. The ratings of the L2 group are largely unaffected and the difference between the two conditions is still clearly visible at 0.69. In the L2 group the ratings have gone down towards the acceptable end of the scale and the numerical difference between the two conditions has increased to 0.27. Post-hoc t-tests showed that the difference in the L2 group is now marginally significant ($t_1(38)= -1.98$, $p=0.055$, $t_2(7)= -4.53$, $p=0.003$)

	DAT > ACC	ACC > DAT
L1 group	1.16 (0.33)	1.75 (0.68)
L2 group	1.31 (0.7)	1.58 (0.85)

Table 5.10 Average acceptability ratings across conditions, SDs in brackets, data points with unrelated corrections removed

One more reduction to the dataset was made after examination of the individual results in the L2 group as the effect of order in the L2 group was much less reliable than in the L1 group. One speaker's data points were removed for exhibiting random rating behavior, judging some of the highly ungrammatical sentences as acceptable and had averages of 3.75 and 4.25 for the accusative-first and dative-first conditions. It could be the case that this participant had misinterpreted the rating scale, especially as the head of the scale with the acceptable/inacceptable endpoints was not repeated on later pages of the task. Unfortunately, this participant did not provide corrections for bad sentences, so it is impossible to know the actual cause of the ratings. Removing this participant from the dataset lowers the averages in the L2 group to 1.22 and 1.52, resulting in a difference of 0.3 between the two ratings, but not affecting the outcome of the ANOVA.

Figures 5.4 and 5.5 below show histograms of the differences between accusative-first and dative-first ratings after the removal of all unwanted data points. A positive value indicates that the accusative-first rating was less acceptable than the dative-first rating and the bigger the value, the bigger the advantage for the dative-first rating. A negative value accordingly indicates an advantage for the accusative-first rating. Ratings that result in a difference of 0 are counted in the bar to the left of the 0.0 on the x-axis.

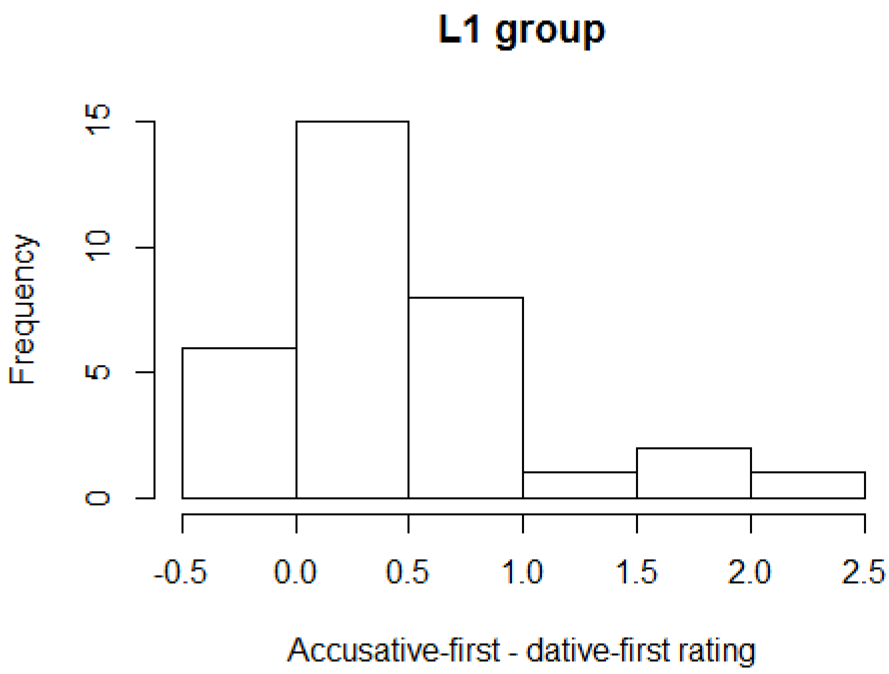

Figure 5.4 Histogram of the acceptability rating differences of the L1 group

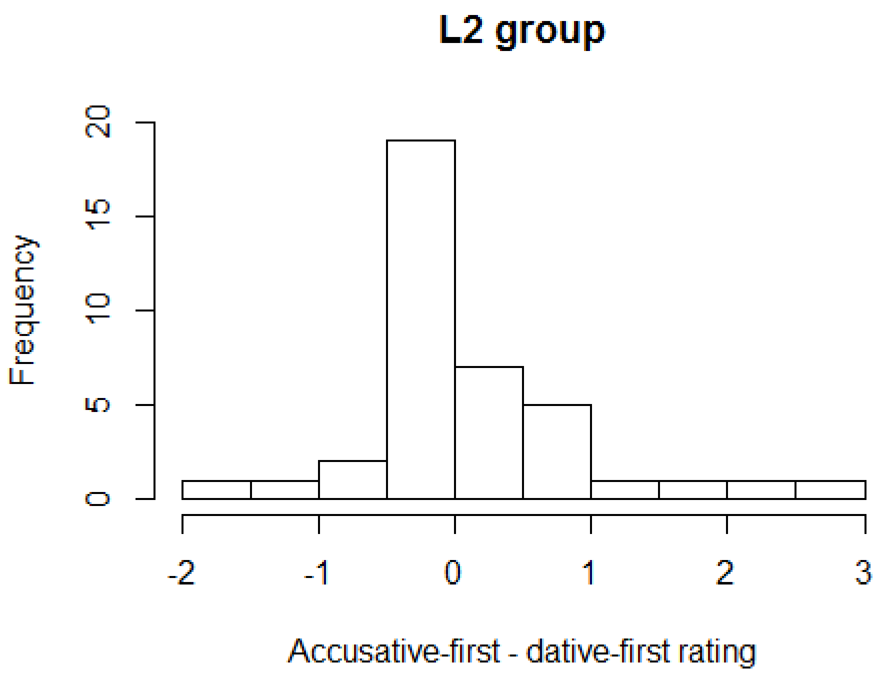

Figure 5.5 Histogram of the acceptability rating differences of the L2 group

It is immediately visible from the histograms that the L2 group shows a much wider spread in rating differences than the L1 group. Only three speakers in the L1 group show a numeric advantage for the accusative-first order and another three speakers in the L1 group showed no preference for either order (both orders were rated with a perfect 1.0). The remaining 27 participants showed a numeric preference for the dative-first order of up to 2.25 points with the majority ranging between 0.25 and 1. The comparatively small skew is also reflected in the comparison of mean 0.58 and median 0.5 that do not differ much. The higher mean can be explained by the differences beyond 2.0. As already indicated by the higher standard deviations in the previous tables and visible in Figure 14b, the L2 group shows a much broader range in rating differences than the L1 group. More L2 participants tend to show an advantage for the accusative-first order, and to a more extreme extent than in the L1 group. This advantage was present in six participants and ranged from 0.33 to 1.67. The abovementioned speaker with the possibly incorrect use of the rating scale was part of this group. The other five speakers all came from different L1 languages (Czech, Macedonian, Polish, Russian, Slovenian). Seventeen L2 participants rated both orders as equally acceptable, usually with a perfect 1.0 rating. This group is also much bigger than in the L1 group. The remaining 16 participants showed an advantage for the dative-first order that ranged between 0.25 and 3.00. The two participants with the most extreme differences rated the accusative-first order at 3.67 and 4.00 exceeding the maximum rating for accusative-first sentences of any participant in the L1 group and judging these constructions as almost ungrammatical. The comparison of mean 0.27 and median 0 illustrates an important point. The removal of item with corrections unrelated to the experimental manipulation leads to a marginally significant difference between the two ratings. However, a large number of participants in the L2 group do not reliably distinguish the acceptability of the two word orders, and the group effect is likely caused by the extreme ratings of participants that consider the accusative-first order as near-ungrammatical.

In order to explore this possibility, I re-ran the analysis of the reduced dataset with items and participants removed as previously described. Additionally, I removed the two L2 participants that had given unacceptable ratings to the accusative-first sentences. This procedure left 518 remaining data points and resulted in ratings of 1.23 (SD: 0.5) in the dative-first condition and 1.39 (SD: 0.56) in the accusative-first condition, a difference of 0.16 that is smaller than in the previous analysis. The between-groups ANOVA still showed the main effect of Order ($F_1(1,67)=23.7$, $p<0.001$, $F_2(1,7)=9.38$, $p=0.018$), and the Group x Order interaction also reappeared ($F_1(1,67)=7.94$, $p=0.006$, $F_2(1,7)=10.85$, $p=0.013$). The by-items ANOVA also showed a main effect of Group that had been absent in the by-subjects ANOVA ($F_1(1,64)=1.88$, $p=0.18$, $F_2(1,7)=6.4$, $p=0.04$). Post-hoc t-tests showed that the marginally significant effect in the by-subject analysis had disappeared: ($t_1(32)= -1.54$, $p=0.13$, $t_2(7)= -3.21$, $p=0.015$).

The L2 group showed more variance in their rating differences than the L1 group (see Figure 5.5 above) and also exhibited some variation with regard to the age of acquisition and how long they had learned German. Especially longer exposure to German could lead to a higher familiarity with the non-canonical word order, as it might have been encountered more often over time leading to a better assessment of the small difference in acceptability between the canonical and non-canonical order. Separate linear regression analyses using L2-specific predictor variables such as proficiency, AoA and years spent learning German were run on the difference in acceptability ratings for the entire L2 group (N=39). Neither of the post-hoc analyses yielded a significant result. An overview of all analyses conducted with the corresponding plots can be found in Appendix B.

The four participants that had been flagged as possible heritage speakers due to their early age of arrival in Germany did not show a uniform rating behavior. Participant 207, 223 and 235 showed ratings that are within the range of the majority of L1 ratings, while participant 229 also showed no difference between the ratings which was more typical for the L2 group than the L1 group. As the ratings of L1 and L2 group still largely overlap, this particular subgroup does not set itself apart from either group.

ID	**dative-first**	**accusative-first**	**difference (accusative – dative)**
207	1.25	1.67	0.42
223	1.25	2.50	1.25
229	1.00	1.00	0.00
235	1.00	2.00	1.00

Table 5.11 Acceptability ratings of L2 participants flagged as potential heritage speakers

Discussion

L1 and L2 speakers did differ in their word order preferences for ditransitive sentences in out-of-the-blue situations. Both groups judged either order as grammatical, but only the native group showed a reliable effect of order.

The ratings of the L1 group were in line with prediction A above as the dative-first order was judged as more acceptable than the accusative-first order. This preference was also expressed in the corrections and comments the participants made, often changing an accusative-first to a dative-first order, while still judging the accusative-first order as acceptable overall. While the difference between the two conditions was small at around 0.6, it was significant and is in line with previous results in the literature that also found a rating advantage for the dative-first order (e.g. Keller, 2000). As pointed out in the predictions, frequency, focus, animacy and case ordering preferences all favor the dative-first order. With the present design it is not possible to disentangle these factors in order to determine whether the source of the preference for the canonical word order lies in either of these ordering constraints, or whether it reflects an actual difference in the underlying syntactic representation.

The overall result of the L2 group showed no significant difference in acceptability between the two orders. Some post-hoc analyses that excluded data points based on unrelated corrections did show marginal effects, but these were likely caused by speakers that made unusually large distinctions in acceptability ratings between the two orders. It is unclear why these participants perceived the accusative-first order as unacceptable. The L2 group as a whole showed little effects of word order frequency in naturally occurring language and the majority of L2 speakers was able to accurately accept a word order that they had little exposure to based on the low frequency of accusative-first orders found in corpus studies (Keller, 2000; Kempen & Harbusch, 2005). The absence of an effect of experimental manipulation in the L2 group could be explained by several factors. One could be an insensitivity to any of the ordering preferences that favor the canonical dative-first order. However, a general insensitivity to ordering factors seems unlikely given that previous studies on the English dative alternation had shown partial sensitivity to the given > new principle in L2 speakers (Park, 2011). A selective inability to use case as an ordering principle would still have left the principle of animate > inanimate to highlight the higher acceptability of the DAT > ACC order.

Another possible explanation could be a task effect as participants in the L1 and the L2 group might interpret the rating task differently. The L1 group seems to have interpreted acceptability as going beyond the binary grammatical/ungrammatical distinction and that out of two fully grammatical sentences one can be more acceptable than the other. The L2 group seems to have focused more on grammatical errors and sentence content. As the two experimental conditions did not differ regarding their grammaticality or their semantic content, both would be assessed as highly acceptable. The high prevalence in the L2 group of corrections that were not related to the experimental manipulation might be indicative of this perception of the rating task as did their performance on filler sentences containing ungrammatical and semantically anomalous constructions. Future research would have to tease apart these factors to better determine their individual contribution to the perception of acceptability by L2 speakers including experiments assessing their ability to use case marking and to determine the context-appropriateness of non-canonical orders.

The information structure of the sentences tested contained decontextualized and therefore all new information. This is usually rendered as a dative-first order in German and most of the Slavic L1s according to descriptive grammars, the exceptions being Macedonian and Bulgarian which favor accusative-first orders for full NP objects. Transfer of the dative-first preference from the remaining L1 backgrounds should have resulted in a more pronounced difference between the two orders based on the predictions of descriptive grammar. However, the small difference in acceptability is not unique in the literature. In her experiments on the elicited production of ditransitive structures in Russian, Kallestinova (2007) found that participants were equally likely to produce IO > DO and DO > IO orders in context free sentences. Preferences for one order over the other only emerged in sentences focusing either of the two objects. This could suggest that the preference for dative-first orders in wide-focus contexts in the majority of Slavic languages as proposed in descriptive grammar is exaggerated, and that actual production frequencies and processing patterns need to be established based on experimental evidence. On a side note, of the three speakers with Bulgarian or Macedonian as their L1 only one had a preference for the accusative-first order that would support a transfer-based account. The other two speakers showed a preference for the dative-first order that went against the preferences in their L1.

This analysis of the offline acceptability rating data showed the limits of the standard method of elicitation of acceptability ratings. Usually, only the absolute ratings are collected without any possibility for the participant to comment on their ratings. The experimenter is then left with the raw data and cannot be sure that the outcome of the rating was actually influenced by the experimental manipulation or by some other unrelated source. In this experiment, participants were encouraged, but not forced, to indicate the source of error for bad ratings and these comments proved very valuable as they allowed the removal of trials in which an unacceptable rating was unrelated to the actual manipulation. In other methods like self-paced reading or various types of priming, comprehension questions or lexical decisions are used to eliminate incorrect trials from the analysis. Asking for a clarification of the judgment could be a method for judgment tasks to eliminate uninformative trials. The thinking-out-loud paradigm has been used for similar purposes and might have strong effects on the judgment task itself as it requires the actual elaboration of the steps taken to reach a certain interpretation. Just asking the participant to mark the source of an unacceptable judgment could show milder task effects as participants do not have to be able to state a fixed grammatical rule to indicate the source of error. The benefits of collecting this additional data are obvious, as it not only allows the elimination of uninformative answers, but gives additional insight into the participant’s thought process.

Summarizing the results of the acceptability rating task, it can be said that German native speakers showed a reliable preference for the DAT > ACC order in ditransitive sentences without context. The mixed Slavic L2 group did not show a preference for one word order over the other and instead rated both orders as equally acceptable. The following SPR task investigates whether the Dative constraint can be applied by L1 and L2 speakers during their online processing, by-passing possible task effects of the acceptability rating task.

5.3 Experiment 2b: Self-paced reading task

Participants

The same participants were tested as described under 5.2. No participants were removed due to their performance in the acceptability rating task. The possibly wrongly assigned rating scale has no influence on the reading behavior and the participants with unusually high average ratings still considered some of the sentences grammatical suggesting no fundamental problem with either word order, especially in online processing. The L1 group therefore consisted of 33 participants and the L2 group of 39 participants.

Materials

The study employed the same design as the offline task with the factor object order and its two levels (dative-first vs. accusative first). The materials were based on the materials used by van de Koot, Felser & Sato (in preparation) in their ongoing AHRC funded project on scrambling in German using cross-modal priming. Sixteen of their items were used in this experiment. As their materials had been created for experiments with native speakers of German some modifications were made to simplify the sentences. These modifications mainly concerned words that were deemed as too difficult or unknown to L2 speakers.

Each item consisted of one or two lead-in sentences that provided a context for the experimental sentence. The context did introduce the accusative object of the critical sentence as the introduction of two unknown objects especially in the scrambled experimental sentence would have been pragmatically awkward after providing the reader with a context. The context was not intended to motivate scrambling per se as it did not suggest as contrastive interpretation. This would have required the introduction of context as a factor with match/mismatch design. As it was not the objective of this experiment to test for context effects, but rather to investigate the application of the Dative constraint, the context was not biasing towards the scrambled structure.

All object NPs in the experimental sentences were either masculine or neuter to avoid ambiguous case marking on the article. The objects had also been matched for length and frequency in the original materials, but as only a subset of the original items was used I recalculated type length and type and lemma log-frequency based on the dlex database (Heister et al., 2011). The exact numbers for the individual items can be found in Appendix B and a summary of the means is shown in Table 5.12.

	Length	**Type frequency**	**Lemma frequency**
Dative	8.875 (3.26)	0.66 (0.88)	0.826 (0.82)
Accusative	7.063 (2.02)	1.04 (0.9)	1.313 (0.82)

Table 5.12 Mean length, type and lemma frequency per case type, SDs in brackets

Note that for one item (*Hochzeitsplaner*, wedding planner) no frequency information could be found in the database. Log frequency was entered as 0 in the following matching procedure. Two sample t-tests showed that neither of these differences is statistically significant (p=0.07 for the t-test on length, p=0.2 for the t-test on type frequency and p=0.1 for the t-test on lemma frequency) and the object NPs can still be considered matched according to these criteria.

The experimental sentences consisted of two coordinated clauses, one main clause ranging between two to five words and a dependent clause that always had the same length of nine words. The dependent clause began with a connector *dass* ‘that’ followed by the subject and the two objects. Depending on the condition, either the dative object came first (dative-first condition) or the accusative object (accusative-first condition). The dependent clause ended with a participle (n=9) or an infinitive (n=7) and an auxiliary. Six verbs were used twice, once as a participle, once as an infinitive. The remaining four verbs were either used in their participle or infinitival form. The tenses employed in the experimental sentences were past perfect, preterite or future tense.

Lead-in sentence

(103) Seit dem Tod ihrer Mutter hatte Eva oft die Schule geschwänzt. Irgendwann kam Zuhause ein Brief von der Schule an.
‘Since the death of her mother Eva had often skipped school. Eventually a letter from school arrived at home.’

Dative-first

(103a) Eva glaubte, dass der Lehrer dem Vater den Brief geschrieben hatte.
Eva believed that the$_{NOM}$ teacher the$_{DAT}$ father the$_{ACC}$ letter written had
‘Eva believed that the teacher had written the father the letter.’

Accusative-first

(103b) Eva glaubte, dass der Lehrer den Brief dem Vater geschrieben hatte.
Eva believed that the$_{NOM}$ teacher the$_{ACC}$ letter the$_{DAT}$ father written had
'Eva believed that the teacher had written the letter to the father.'

To make sure that the participants in the L2 group were familiar with the nouns used in this task, all nouns that were not listed as belonging to the word list of the B2 level language certificate *Zertifikat Deutsch* by the Duden webpage (www.duden.de) were included in a vocabulary list that was given to the participants after the completion of the self-paced reading task. Participants were asked to indicate any unfamiliar words on the list.

During the SPR task participants were also asked eight comprehension questions that probed their interpretation of the previous sentences corresponding to one question after half of the items. Due to a design flaw, the expected answers were not counterbalanced. There were five questions expecting a negative answer and three expecting a positive answer. For the questions that expected a negative answer, the comprehension question asked about a different noun than either the subject or the object of the experimental sentence. Questions expecting a positive answer dropped one argument in order not to repeat the experimental sentence verbatim.

Each experimental list began with four unrelated practice items and one practice comprehension question. There were 43 filler sentences, 24 of which were items for the German experiment on particle verbs (see section 8.2), the other 19 of which were actual fillers. Of these fillers six were pseudofillers reused from the materials of the original crossmodal priming task. Seven fillers were pseudofillers modeled after items from Experiment 4b and included German particle verbs that had not been split in a V2 context, resulting in ungrammatical sentences. The remaining six fillers contained prefixed verbs in the lead-in sentences and the corresponding bare verbs in the experimental sentences. There were no comprehension questions for the fillers. Altogether, the participants saw 63 sentences and 25 comprehension questions including practice items. Sixteen experimental sentences and eight comprehension questions of these belonged to Experiment 2b. The items were pseudorandomized using the software provided by www.random.org/lists such that there were eight comprehension questions in each of the three presentation blocks containing 20 sentences. Each block contained a balanced number of sentences from each condition. The odd number of 63 sentences results from the 19 fillers that were used and did not affect the Latin Square design as the fillers were the same for all participants. The altogether 40 experimental sentences of experiments 2b and 4b were spread across eight presentation lists in regular Latin Square design. As there were only two conditions in experiment 2b, half of the lists contained the same item. There were also special lists for left-handed participants that included a reversal of the answer buttons for the comprehension questions.

Procedure

The SPR task was the central part of the experimental session outlined in Section 5.2 and was programmed in the same way as the SPR task described in Section 4.4. Instructions again told them to be as quick and accurate as possible in their reading. In addition, participants were informed that they would encounter some sentences that might seem awkward or ungrammatical and that this was not an error, but intentional. Participants had the possibility to take two breaks during the experiment, after one and two thirds of the items respectively. The self-paced reading task took 10-20 minutes for the L1 group and 20-30 minutes for the L2 group.

Predictions

A – If the more frequent word order is processed faster than the less frequent order, then the DAT > ACC condition should generally elicit faster reading times as this is the predominant object order in German ditransitive sentences.

B – If Müller's (1999) proposal is correct in assuming ACC > DAT as the canonical word order in ditransitive sentences, then no reanalysis should take place and the ACC > DAT order could show faster reading times than the more frequent DAT > ACC order that would require a reanalysis.

L2 predictions

Different approaches towards the ordering of objects in ditransitive sentences in the Slavic native languages of the L2 group have been reviewed in Section 5.1. More specific predictions for the L2 group’s online performance will be derived from these approaches in addition to the more general predictions above.

C – In the literature on word order variations in the Slavic languages, the given/new or topic/comment principle is the predominant ordering principle. If the simple, non-contrastive mentioning of the accusative object in the context is enough motivation for the L2 speakers to consider it given information, they should show faster reading times for the ACC > DAT order than for the DAT > ACC order as the former adheres to the given > new structure, while the latter does not (see Slioussar, 2011). I do not expect the same effect in the L1 group as the given/new principle in German is only one of many ordering principles suggested in the literature and likely plays a weaker role than in the Slavic languages.
D - If non-native speakers do not use case information to immediately assign thematic roles and delay their assignment until they reach the lexical verb, any effects for the non-native group could be found at the lexical verb, not earlier.

Results

Comprehension questions

Before getting into the reading time data of Experiment 2b, I will present an analysis of the comprehension questions as low performance on the comprehension questions is a criterion for exclusion from the analysis of the reading times. The question from the practice session was excluded from this analysis leaving 24 comprehension questions. Eight of these questions belonged to items from Experiment 2b and 16 to items from experiment 4b corresponding to questions after 50% and 67% of all experimental items respectively. Table 5.13 below shows the overall accuracy scores and the accuracy scores for each experiment separately.

A first look at the overall accuracies and also the accuracies for each experiment shows that unlike in the Norwegian experiments, there is no difference between questions targeting items from Experiments 2b and 4b and the overall accuracies are slightly higher: 75.9% compared to 71.3%. A between-groups ANOVA with experiment as a within factor showed no effects of Group ($F_1(1,70)=2.37$, $p=0.13$) or Experiment ($F<1$), nor a Group x Experiment interaction ($F_1(1,70)=2.81$, $p=0.98$). The majority of participants in both groups seems to have paid attention to the task and were able to complete it. But there were exceptions that are visible in the accuracy ranges reported in Table 5.13. One

L2 participant had an overall accuracy of 37.5% and also contributed the lowest accuracy scores for each individual experiment. Another three participants (one from the L1 group, two from the L2 group) had overall accuracy scores at 50%. Their performance could be attributed to guessing either due to lack of attention or due to insufficient comprehension of the previous sentence. An additional participant from the L1 group had an accuracy score of 37.5% for the items of Experiment 2b, even though her overall accuracy was 62.5%. These 37.5% set her apart from the rest of the remaining L1 group and suggest a problem with either the construction or the questions. The data of these five participants will be excluded from the subsequent reading data analysis.

	L1 group (N=33)	**L2 group (N=39)**	**Both groups**
Overall	79.04 % (0.11) (range: 50-100%)	73.29 % (0.12) (range: 37.5-95.8%)	75.9 % (0.12)
Experiment 2b	76.89 % (0.19) (range: 37.5-100%)	75.64 % (0.17) (range: 25-100%)	76.2 % (0.18)
Experiment 4b	80.11 % (0.11) (range: 56.3-100%)	72.12 % (0.14) (range: 43.8-93.8%)	75.78 % (0.13)

Table 5.13 Mean accuracy scores for comprehension questions of Study 2 and 4, SDs in brackets

Data cleaning procedure – self-paced reading data

The original dataset contained data from 33 participants in the L1 group and 39 participants in the L2 group resulting in 1152 experimental trials. The five participants with low overall accuracy scores (80 trials, 6.94%) were then removed leaving 1072 experimental trials. In a next step, all trials containing unknown vocabulary as indicated in the vocabulary sheet by the L2 group were removed (15 sentences, 1.3%). The dataset that went into the following analysis contained 1057 sentences. As the main sentences varied in length, only the embedded sentences beginning with the connector *dass* 'that' will be analyzed, the remaining data is dropped. This leaves nine segments per sentence that were analyzed. The main regions of interest are the four segments that were manipulated containing the first object article (Art1), the first object noun (N1), the second object article (Art2) and the second object noun (N2). Additionally, the main verb was analyzed as thematic role assignment is finalized at this point. It also served as a spillover region to capture any delayed effects. In preparation for the main analysis, extreme values and outliers were removed for each segment separately as the length of the segments varied and especially the sentence-final segment containing the auxiliary would have disproportionately affected the exclusion criteria due to sentence wrap-up effects causing highly elevated reading times. Extreme values were identified by visual inspection using a histogram of the raw reading times. Outliers were removed based on a +/- 2.5 SD range around the participant mean per segment. A table with the cutoff points for extreme values and number of outliers removed for each segment can be found in appendix B. Between 25 and 34 data points were removed from each segment corresponding to no more than 3.22% of the data points.

Main analysis

Table 5.14 below lists the mean reading times of the embedded clause across conditions for both participant groups separately after the data cleaning procedure described above had been applied. The L2 group had generally slower reading times than the L1 group in each segment and I therefore expect main effects of Group in all subsequent analyses.

A full graphical depiction of the reading times across the entire embedded sentence can be found in appendix B. Figures 5.6 and 5.7 below display a focus on the region of manipulation plus the spillover region (main verb).

Table 5.15 and Figures 5.6 and 5.7 show reading times differences between the two conditions at the object nouns in both groups. As the objects had been matched for length, this difference has to be due to the experimental manipulation. The position of the object noun within the sentences also seems to play a role as dative and accusative objects both showed higher reading times when they occurred as the second object. In addition to these effects directly on the object noun, the L1 group additionally shows reading differences at the second object article and the main verb.

	dass	**Art0**	**Subj**	**Art1**	**N1**	**Art2**	**N2**	**Main verb**	**Aux**
L1 group									
Dative-first	354 (69.98)	312 (64.71)	336 (98.29)	331 (88.47)	364 (91.77)	351 (71.93)	357 (79.79)	401 (88.32)	575 (270.73)
Accusative-first	367 (72.34)	315 (70.3)	336 (90)	335 (91.94)	348 (90.03)	337 (76.87)	381 (111.3)	431 (110.77)	546 (254.29)
L2 group									
Dative-first	446 (129.39)	410 (176.34)	518 (277.19)	430 (151.46)	708 (388.58)	472 (182.6)	624 (266.46)	767 (325)	1000 (466.9)
Accusative-first	466 (159.39)	416 (161.41)	509 (253.5)	433 (160.18)	566 (255)	465 (191.58)	753 (379.17)	793 (355.55)	949 (501.7)

Table 5.14 Mean reading times per segment across conditions. SDs given in brackets, region of manipulation shaded

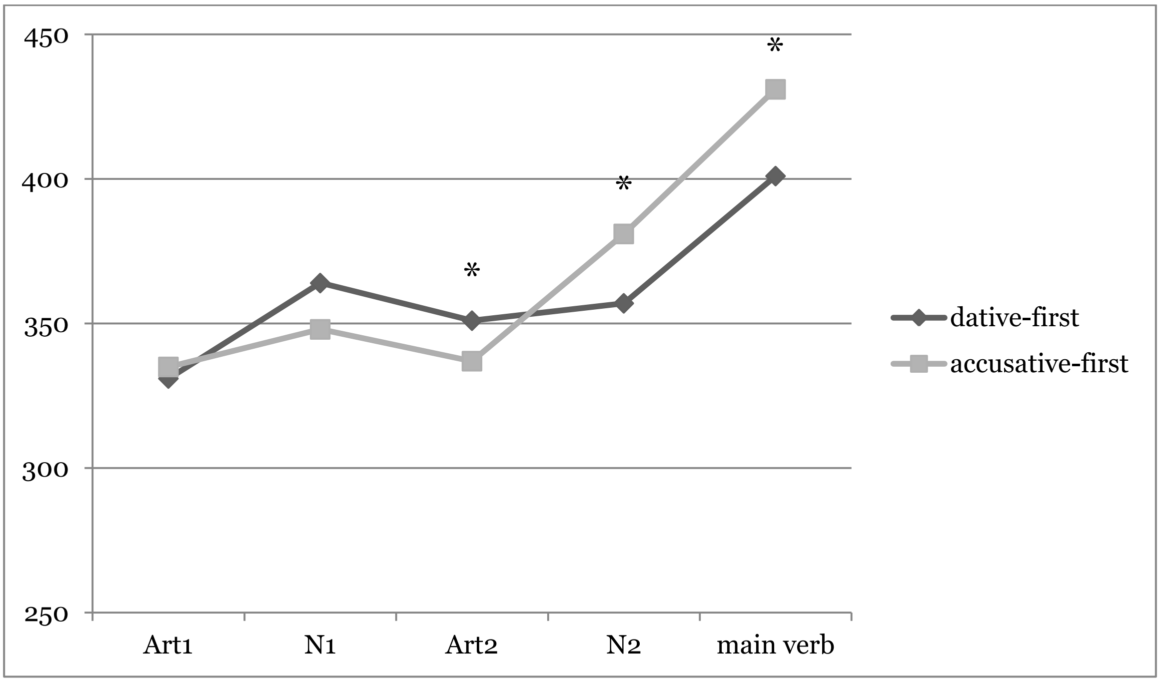

Figure 5.6 Reading times at region of manipulation + spillover region (L1 group)

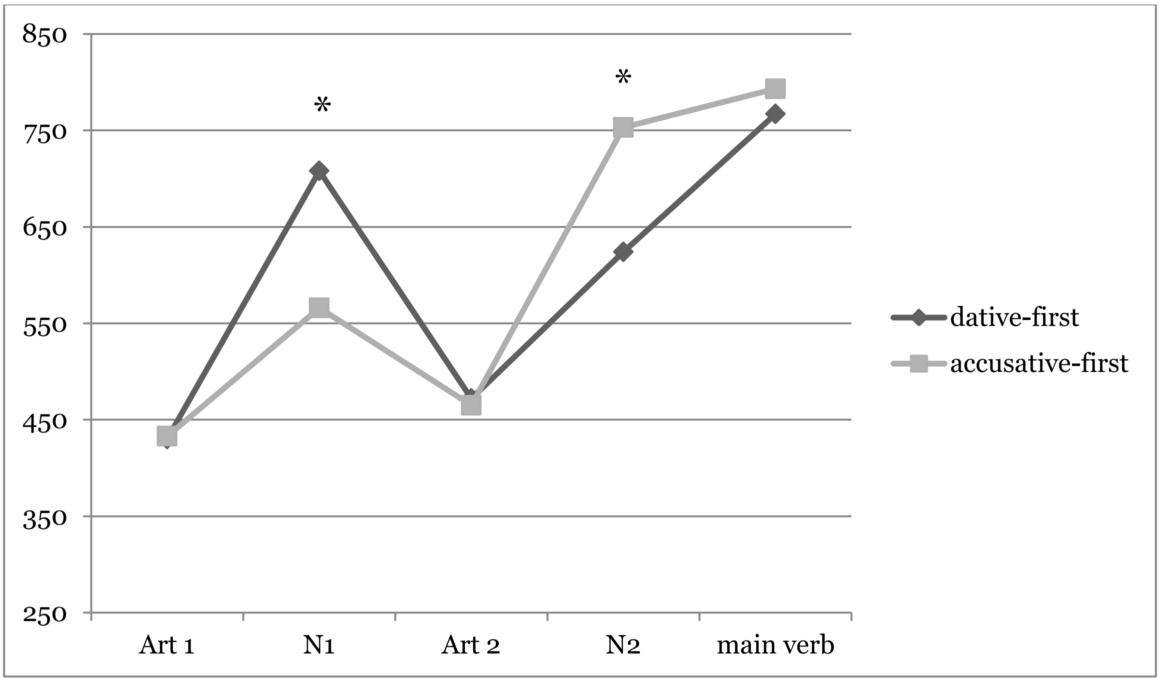

Figure 5.7 Reading times at region of manipulation + spillover region (L2 group)

The box-cox power transformation suggested a reciprocal square root transformation to improve normality in eight out of nine segments. In order to apply the same transformation to all segments and due to the fact that the remaining ninth segment was not part of the manipulation, the reciprocal square root transformation was applied to all nine segments. The between-groups

ANOVAs showed a main effect of Group for every single segment analyzed, reflecting as already expected the slower reading times in the L2 group. Prior to the region of manipulation, a main effect of Order was found at the connector ($F_1(1,66)=6.7$, $p=0.012$, $F_2(1,15)=7.22$, $p=0.017$). This difference between the two conditions before the actual manipulation then disappears in the following two segments. There was no significant effect of Order at the first object article (Art1), but a significant effect of Order ($F_1(1,66)=22.44$, $p<0.001$, $F_2(1,15)=9.7$, $p=0.007$) and a Group x Order interaction ($F_1(1,66)=5.9$, $p=0.018$, $F_2(1,15)=4.6$, $p=0.048$) at the first object noun (N1). Every segment following the first object noun showed a main effect of Order. The full between-groups ANOVA results can be found in appendix B. Based on the Group x Order interaction at segment N1, I analyzed the two groups separately for the remaining segments in order to apply the same analysis to all segments and not to jump from between-groups to by-groups and back to between-groups analyses.

Paired t-tests run on the N1 section showed no effect of Order in the L1 group ($t_1(30)= -1.53$, $p=0.14$, $t_2(15)= -1.08$, $p=0.3$) and a significant effect in the L2 group ($t_1(35)= -4.89$, $p<0.001$, $t_2(15)= -3.3$, $p=0.005$), reflecting longer reading times in the dative-first condition and the Group x Order interaction. In the Art2 section, there was an effect of Order ($t_1(30)= -2.4$, $p=0.02$, $t_2(15)=-1.55$, $p=0.14$) in the L1 group, but not in the L2 group ($t_1(35)= -1.2$, $p=0.24$, $t_2(15)= -0.99$, $p=0.34$). This reflects the slower reading times for the dative-first condition at this point in the L1 group and no difference in reading times in the L2 group. It was not supported by a Group x Order interaction, though. In the N2 section, I found significant effects of Order in both groups, but only in the by-subjects analysis. L1 group: ($t_1(30)=2.49$, $p=0.02$, $t_2(15)= 1.88$, $p=0.08$), L2 group: ($t_1(35)= 4.22$, $p<0.001$, $t_2(15)= 1.95$, $p=0.07$). Both groups show elevated reading times in the accusative-first condition, while reading the dative object at this point.

Outside the region of manipulation, the main verb in the spillover region is of interest due to the assignment of thematic roles and the final confirmation that this embedded clause is a grammatical ditransitive clause. Paired t-tests at the main verb showed an effect of Order for the L1 group ($t_1(30)=2.15$, $p=0.04$, $t_2(15)=2.06$, $p=0.057$), but not in the L2 group ($t_1(35)= 1.04$, $p=0.3$, $t_2(15)=1.96$, $p=0.069$). This result reflects the longer reading times for the accusative-first condition in the L1 group, and no reliable reading time differences in the L2 group. The sentence-final segment showed a significant effect of Order in the L1 group ($t_1(30)=-2.75$, $p=0.01$, $t_2(15)=-2.5$, $p=0.025$) and a marginally significant effect in the L2 group ($t_1(35)=-1.76$, $p=0.09$, $t_2(15)=-2.04$, $p=0.06$). Both results go in the same direction with slower reading times in the dative-first condition, but as this is the end of the sentence, no strong conclusions should be drawn from this section.

Summarizing the statistical analysis, it has to be observed that effects of slower reading times are limited to the object nouns in the L2 group with longer reading times for the dative object. In the L1 group, significant effects of slower reading times only begin at the second object article, but are found in every following segment. They show slower reading times for the accusative-first condition at the second object noun and the main verb.

Post-hoc analyses

As more experience with German in general could also lead to more experience with non-canonical word orders, I used the biographical data of the participants to run separate post-hoc linear regressions. The analyses used as predictor variables proficiency as measured by the Goethe score, AoA, years spent learning German and years spent living in Germany to investigate whether the reading times between the conditions were affected by any of these variables. I only present the analyses for those segments in which L1 and L2 group differed and any of the abovementioned factors could contribute to a progress towards more native-like behavior. More native-like behavior in this case would mean a reduced difference between reading times at the first object noun and a difference at the main verb. A full overview of the post-hoc analysis for the entire region of manipulation can be found in Appendix B. The only marginally significant result in the entire series of analysis was an influence of Goethe score in the region of the main verb ($t=1.75$, $p=0.089$). With increasing proficiency, participants tended to show faster reading times in the dative-first condition compared to the accusative-first condition, but there were also a number of participants in the higher proficiency range that showed the exact opposite pattern.

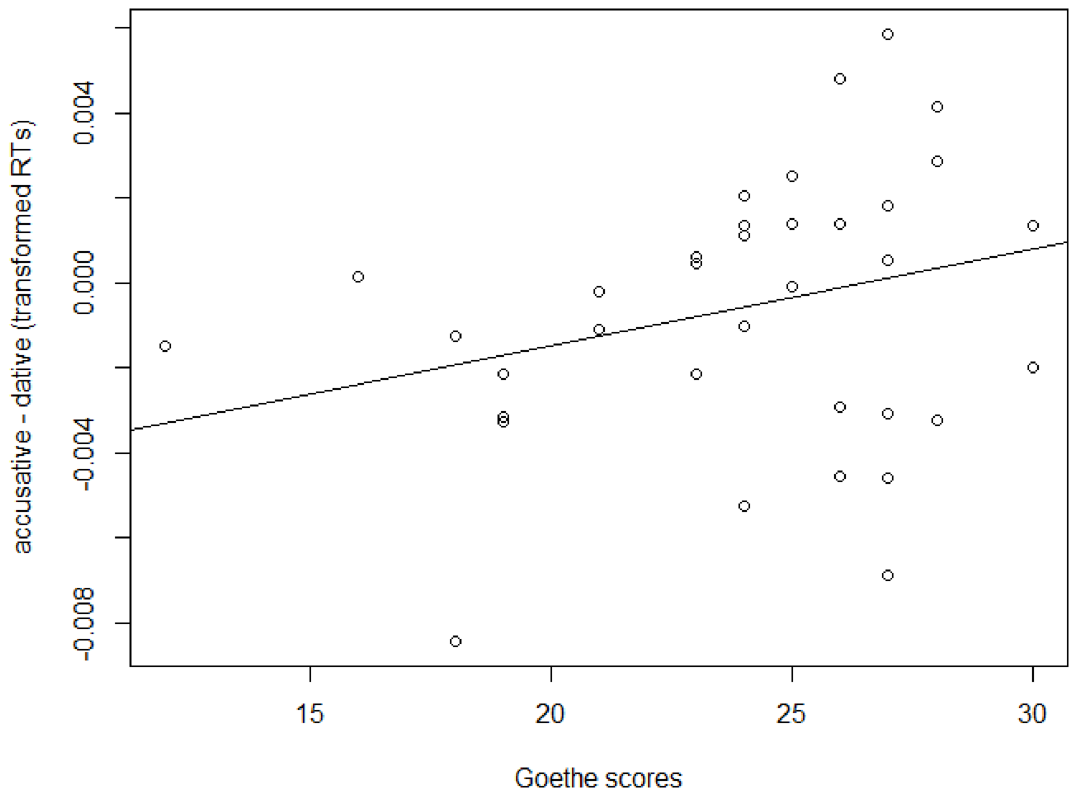

Figure 5.8 Linear regression on the main verb segment with Goethe score as predictor

Discussion

The self-paced reading data revealed different reading time patterns for the L1 and the L2 group. The reading time pattern of the L1 group corresponds roughly to prediction A that assumed longer reading times for the less frequent non-canonical word order. The L1 group had significantly longer reading times for the non-canonical accusative-first condition at the second object NP (N2) and at the main verb. The result of the L2 group does not reflect any of the predictions above. This group only showed significantly longer reading times for the dative noun and this slowdown was not related to the order manipulation.

The German L1 group showed effects of the order manipulation at three points of the experimental sentence. The first effect was found on the second object determiner with faster reading times in the accusative-first condition. This means at this point the participants took longer to read the accusative determiner *den* than the dative determiner *dem*. This result was only significant in the by-subjects analysis and not in the by-items analysis. It is difficult to interpret as the numerical difference between the conditions in the preceding segment, the first object NP, was similar, but did not turn out significant due to higher variance. The significant result for the determiner could be simply due to spillover from the dative object. The dative object at the first object position had numerically higher reading times than the accusative object and although this did not result in a significant difference, it could hint at a slightly higher processing effort for the dative compared to the accusative object that continues into the next segment. Whether this higher processing effort for the dative is caused by the nature of the dative case itself (form, frequency etc.) or by the fact that the dative object was new information that had not been introduced in the context cannot be teased apart at this point. A second possibility is that participants expected an indefinite determiner at this point as Pappert et al (2007) had found a higher occurrence of definite > indefinite orders in dative before accusative sentences both in their corpus data and in their completion questionnaires. The expectation of the speakers was not met, causing higher reading times.

The next significant effect was found in the following segment, the second object noun. There, L1 speakers showed higher reading times for the dative object (the non-canonical accusative-first condition) than for the accusative object (the canonical dative-first condition). This could be caused by the reinterpretation of the sentence as a ditransitive abandoning a previously possible transitive interpretation or a more general effect of the new information contained in the dative NP. However, the numerical difference between the two conditions is bigger than at the first object noun (N1). The variance in reading times for the dative object is also considerably higher than for the accusative object at N2, while they were comparable at N1. This suggests that the disadvantage for the dative object is further aggravated by the non-canonical position of the dative. The third significant effect of order was found on the main verb and went in the same direction as the effect in the preceding section with longer reading times in the non-canonical accusative-first condition. These latter two effects are in line with a frequency-based explanation, as the DAT > ACC order is by far the more frequent order in written German (Kempen & Harbusch, 2005). There is no evidence to support claims such as Müller's (1999) that assumes the ACC > DAT order to be the canonical word order that should be processed more quickly than a non-canonical DAT > ACC order in his view. Aside from being the canonical structure for ditransitive sentences, the DAT > ACC order also complies with two ordering constraints: the Dative constraint that was manipulated in this experiment and the Animacy constraint that was present as a confounding factor. If we take Primus' (1998) approach that assumes a strong association of thematic role and animacy this confound is almost unavoidable and the two constraints might not exist independently from each other. In fact, both constraints have been found to be strongly associated and equally influential in corpus and sentence completion data by Pappert et al (2007). Native speakers seem to be very aware of the multiple small differences between the two object orders in ditransitive sentences and the association of several of these factors (frequency, canonicity, compliance with ordering constraints) in the DAT > ACC order ensures smoother processing compared to the ACC > DAT order. Further studies possibly involving verbs with an ACC > DAT ordering bias could tease apart the effects of each of these factors in native processing.

The L2 group did not show an effect of order on the main verb and only showed significant reading time differences on the object nouns themselves. This difference did however not reflect the order manipulation as the dative noun was always read more slowly than the accusative noun. This pattern does not reflect any of the predictions above as the reading times differences found were not a result of the experimental manipulation.

The slower reading time effect for previously unmentioned dative objects was partly also found in the native group, but there it was only significant at the second object noun, while the effect appeared on both object noun positions in the non-native group. It is likely that this effect is caused by the appearance of new information provided by the dative compared to the old information of the accusative. There is no clear evidence that the violation of the given > new principle in the dative-first condition added onto the already existing reading time disadvantage for the dative. The numerical disadvantage for the dative is bigger in the dative-first condition (142 ms) than in the accusative-first condition (129 ms), but this 13 ms difference is not significant and not comparable to the values found in native-language experiments (e.g. Slioussar, 2011 for Russian) when the given > new principle was violated. In the absence of an experimental manipulation of context as had been done by Slioussar, it remains unclear whether the disadvantage of the dative is caused by the information structure or reflects possible general problems with the dative case. The fact that no effect was found on the object determiners could suggest that either the case information is not used at all or its processing is delayed until the appearance of the object noun. The overall null result with regard to L2 speakers' processing of canonical vs. non-canonical orders in this experiment is a bit surprising based on the established results in the literature. In previous studies investigating the online processing of non-canonical orders such as object-first sentences, L2 speakers had been found to be sensitive to the canonicity manipulation by showing elevated reading times for the non-canonical order and lower accuracy scores in comprehension questions. L2 speakers were found to be able to use disambiguation cues that signaled non-canonical word orders (see e.g. Gerth et al., 2015, the literature review in Section 3.2 and also my own results from study 1). What makes this experiment different from the previous studies that could be responsible for the absence of an effect of canonicity in the L2 group? Previous studies had found results for subject/object ambiguities that were clearly disambiguated by cues such as verb agreement or case marking and usually involved abandoning the preferable interpretation of an ambiguously marked noun in favor of a less preferable interpretation. In this experiment there is no ambiguity with regard to case assignment, especially for the objects the order of which was manipulated, so there is no additional disambiguation needed. The only possible ambiguity is in the accusative-first condition when it is not clear at the first object noun whether the sentence is transitive or ditransitive. The comparison of the strength of ordering constraints by Keller (2000) suggested that the dative constraint that was manipulated in this experiment is much weaker than the nominative constraint that was manipulated in the experiments that investigated L2 speakers. The canonicity manipulation in the present experiment might not have been salient enough for the L2 speakers to cause processing differences as it contained only an unambiguous object/object manipulation instead of an ambiguous subject/object manipulation. L2 speakers might not feel the need to reanalyze the non-canonical sentence in the present experiment as the manipulation does not change the content of the sentence, only its applicability in certain contexts unlike in the studies involving subject/object ambiguities. But the small difference in structures investigated does not give a sufficient explanation of the null result as there were additional factors that strongly favored the dative-first order, such as frequency and noun animacy.

Order frequency should have favored the dative-first order as it is the more frequent order in German, and while all Slavic L1 languages also exhibit the accusative-first order, the dative-first order is the canonical and more frequent order in all Slavic languages but Bulgarian and Macedonian. Slioussar's processing study and Kallestinova's production study on Russian had shown that speakers are sensitive to the change in order. They showed slower reading times when the sentence did not fit the information requirement of the context and adapted their production of ditransitive sentences to the context provided. Only when there was no context at all were both orders, dative-first and accusative-first, produced with equal frequency. Given the lack of experimental evidence in processing as well as production from the other Slavic languages, I have to rely on the assumptions of descriptive grammar that states dative-first as the more frequent order, excluding Bulgarian and Macedonian. As context was not experimentally manipulated in this experiment, it is not possible to estimate the individual effects of order frequency (favoring the canonical dative-first order) and information structure (favoring the accusative-first order). They might unintentionally have canceled each other out leading to the null result and requiring follow-up investigations that either manipulate context or leave it out altogether. More research on the application and ranking of ordering constraints in the various L1s is also needed to be able to make better predictions on which manipulations should be most effective.

Given the problematic nature of case marking for L2 speakers (Clahsen et al., 2010; Papadopoulou et al., 2011; Rankin, 2014), the L2 group might not have been able to use the case marking to assign thematic roles. Participants would then fail to realize that these are two different arguments order (DAT > ACC and ACC > DAT) and instead represent them as animate NP > inanimate NP and inanimate NP > animate NP. Only an independent Animated First Principle (Tomlin, 1986), would then predict an reading time advantage for the animate > inanimate order present in the dative-first condition. The independent existence of this principle has been challenged in the literature and the absence of an effect could support the claim by Primus (1998) against an independent Animated First Principle. Primus assumes that the preference for the animate > inanimate order is closely linked to the association of animacy with thematic roles and should not be considered an independent ordering principle. If participants were not able to use the case marking information on the determiners, they could not reliably assign thematic roles and any ordering preferences based on thematic role assignment would have become void. As unknown vocabulary had been removed from the data analysis, it can be assumed that the L2 participants did not have problems identifying the animacy of the nouns, even if they might have struggled with case.

The possible failure to use case information to assign thematic roles would also have little consequence on the results of the comprehension questions as the exact thematic role assignment was not tested. For negative answers, one of the arguments had been replaced by another noun and a shallow parse would be sufficient as the noun in the question could not be matched with the nouns in the experimental sentence.

L2 specific factors showed little effect on the reading time patterns. The only marginally significant effect was that of proficiency as measured by the Goethe score on the main verb. The trend towards a difference in reading times for more proficient speakers could suggest a change in the underlying processing routine based on the more successful use of case marking with growing proficiency.

No influence of order frequency or of object animacy and overall inconclusive evidence regarding the use of case marking, are in line with approaches to L2 processing that assume qualitatively different processing in L2 speakers as opposed to L1 speakers, like the SSH. If L2 speakers only compute a shallow parse of the sentence without necessarily specifying the thematic roles of the individual objects, both orders are just plain variations of each other that neither differ in their syntactic complexity nor in their focus structure. They only differ in the order of three nouns that have different animacy values if the case information is not or cannot be used. Even a shallow representation that assigns subject and object status would not be able to sufficiently differentiate the two conditions as they would both be represented as subject – object – object. The effects of animacy that have been found to facilitate the L2 processing of subject/object ambiguities (e.g. Jackson, 2007) seem to be less influential in the processing of object-object structures.

Summing up the results of the self-paced reading task, a difference between L1 and L2 speakers in their processing behavior was found. The L1 group showed effects of the order manipulation, especially at the main verb, suggesting a higher processing effort for the non-canonical word order. The L2 group only showed significant reading time differences at the object nouns with generally higher reading times for the dative object. It is unclear whether this effect was caused by processing problems associated with the dative case marking or because the dative noun had not been introduced in the context before. There was no difference in reading times at the main verb.

5.4 Conclusion

The two experimental groups produced the same patterns in the acceptability rating task and the self-paced reading experiment. However, these patterns differed between the two groups. Native speakers showed a general advantage of the dative-first order in acceptability and reading times. The non-native speakers made no difference between the two orders, neither with regard to acceptability nor in processing.

Q2.1 Do L1 and L2 speakers of German perceive one object order as more acceptable than the other in no context situations? Is there evidence for a gradience in acceptability in both groups?

No. Only the L1 speakers showed a statistically significant difference between the two orders in the acceptability ratings. In line with previous research involving the acceptability of word orders in ditransitive sentences, both orders were rated as overall acceptable, but the dative-first order was reliably favored over the accusative-first order. This was additionally supported by the high number of corrections made by the L1 group that changed the accusative-first order to a dative-first order. The L2 speakers perceived both orders as equally acceptable. Their corrections also showed that they paid more attention to content than structure. There were no effects of proficiency or time spent learning German or any other factors that indicate a development towards a gradient acceptability over time.

Q2.2 Are canonical word orders processed more easily than non-canonical orders? Is a processing advantage visible in native and non-native speakers alike?

No. Only the L1 group showed faster reading times for the canonical dative-first order and this effect was present at the second object noun and the main verb. The reading time data of the L2 group failed to show any effects based on the canonicity manipulation. The only effect found could be attributed to the novelty of the dative object as it elicited longer reading times than the accusative object independent of its position in the sentence. It remains unclear whether the L2 group was actually able to use the information provided by the case marking on the determiner to establish thematic relations and recognize the non-canonical order as such.

Q2.3 Is the L2 result possibly influenced by the variance in Age of Acquisition, the time spent learning German and the length of stay in Germany that varied considerably between the participants?

L2 specific factors with regard to proficiency, AoA and time spent learning German played no role in explaining the data pattern neither in the acceptability rating task nor in the self-paced reading data. More proficient learners had faster reading times than less proficient ones, but all other analyses failed to reach significance.

5.5 Intermediate discussion: Objects in non-canonical positions

A comparison of the experiments reported in Chapter 4 and 5 shows two different results regarding the processing of objects in non-canonical positions.

In the object topicalization study reported in Chapter 4, there was no difference between the L1 and the L2 group as both showed effects of the order manipulation. Both had similar accuracy scores in the agent identification task and their reading times also showed a nearly identical pattern. The influence of NP animacy was present in both groups and slightly more pronounced in the L1 group.

In the German midfield scrambling study reported in Chapter 5, there was a difference between the L1 and the L2 group. The L1 group was sensitive to the order manipulation in both tasks with different acceptability ratings and reading times for the canonical and non-canonical object orders. The L2 group did not seem to be sensitive to the order manipulation neither in the offline nor the online task.

The different patterns across experiments could be explained by the nature of the manipulation. Object topicalization, if noticed by the parser, requires a major revision of the syntactic representation, thematic role assignment and sentence prosody that affects the entire sentence as the thematic roles are practically reversed. It was a challenging task for both participant groups. Scrambling on the other hand requires a much smaller revision of the syntactic representation and does not change the content of the sentence nor the thematic role assignment. This makes the task superficially easier. Scrambling is usually motivated by context or information structure requirements that might be too subtle for L2 participants to notice. While the action in an object topicalization can be completely misunderstood if case information is ambiguous or falsely interpreted, the difference in animacy between the two object nouns that is often present in midfield scrambling of objects can save the parser from misinterpreting the action without necessarily using case marking. An underspecified syntactic representation can be detected more easily for object topicalizations than for scrambling.

Case marking might also play a role as the German scrambling sentences required the correct use of case information for successful parsing and the successful detection of an actual difference between the two orders. As L2 speakers tend to struggle with case marking, it is possible that they were not able to efficiently use case marking in the self-paced reading task and could not identify one order as the canonical order and the other as the non-canonical order. Instead, they might have represented both as subject-object-object sentences. The Norwegian topicalization sentences on the other hand did not contain any case marking, instead the pure order of main verb and NP2 had to be used to detect the experimental manipulation which is probably the easier of the two tasks.

The L2 group showed a similar performance to the L1 group only in Study 1. This could suggest that non-canonical structures requiring major revisions of the canonical structure are easier to detect for L2 speakers as being non-canonical than those structures that only involve minor changes. Additionally, having no case marking and no verb agreement to disambiguate canonical and non-canonical sentences in study 1 seemed to be less of a disadvantage, and L2 participants were as successful in using surface word order to identify non-canonical object topicalizations as were L1 participants. Having to rely on case to identify the scrambled order in Study 2 might have put L2 participants at a disadvantage compared to L1 speakers.

The following two experiments on particle verbs will provide more data on optional word orders that involved subtle changes signaled by order differences placing Study 3 between Study 1 and 2. Study 4 on German particle verbs provides new data on how L2 speakers processed verbs that appear in drastically different shapes depending on the syntactic context.

PART II

-

PROCESSING OF PARTICLE VERBS

6 Introduction

This chapter marks the introduction to Part II of this thesis which deals with the processing of verbs with particles in Norwegian and German. Studies 1 and 2 were mainly focused on object movement and the processing of objects in non-canonical positions. Study 3 on Norwegian particle verbs serves as a bridge between Part I and II as it investigates the placement of objects of particle verbs in the Norwegian midfield. The optional placement of the object either before or after the particle is as subtle as the alternations in German scrambling used in Study 2 and surface order is used to signal the difference between the two orders as in Study 1 on Norwegian object topicalizations. The combination of midfield movement with surface word order as disambiguation serves to disentangle the role of either factor in the ability of L2 speakers to perform like native speakers. Study 4 does not involve object movement or non-canonical word orders. Rather, it uses German particle verbs to assess the ability of L2 speakers to correctly use syntax in their processing of these particle verbs as the particle is split from the verb in specific syntactic contexts that will be explained in detail below. Whether the particle is split or not is fairly salient, because if it is split, it creates an additional filler-gap dependency. The task of recognizing whether a particle has been split correctly is difficult, because it involves the knowledge that the verb in question is actually a particle verb and the knowledge of the syntactic rules under which splitting applies. The methodology used in the following two studies is the same as in Study 2: Acceptability ratings to assess the ability of the participants to differentiate between the two orders and a self-paced reading paradigm to investigate their online processing.

Chapter 6 gives a general background on theoretical approaches to particle verbs and reviews the research literature with respect to object and particle shift in Norwegian, and focusing on the processing of discontinuous dependencies generally, and particle verbs more specifically. Section 6.1 reviews the main theoretical discussions surrounding the topic of particle verbs. There is no generally agreed upon term for these verbs and even more problematic, the different approaches do not agree what actually constitutes a proper particle verb. This chapter seeks to provide an overview of the current debate and to clearly outline the terminology and assumptions used in this dissertation. Section 6.2

reviews selected previous studies on particle verbs. Although Norwegian particle shift and particle verbs in German have received a lot of theoretical attention, processing studies in either language are rare as the dominant form of research involving particle verbs are corpus studies, production studies and studies focusing on the derivation of the semantics of particle verbs. The latter type of study is closely related to the theoretical discussion surrounding particle verb status and opaqueness. Particle shift in Norwegian in which the direct object moves across the particle, is often considered a specific type of the more general object shift that involves the movement of pronominal objects across sentential adverbs. Studies on this phenomenon are reported in section 6.2.1. Section 6.2.2 reports studies mainly on English and German particle verbs and also related non-local dependencies.

6.1 Background: Introduction to particle verb theory

In the research on verbs with particles such as *drink up*, German *austrinken* and Norwegian *drikke opp*, terminology is the first issue to address. As an illustration of terminological difficulties, the introductory chapter of the book *Verb-particle explorations* by Dehé, Jackendoff, McIntyre, & Urban (2002) lists **particle verb, verb-particle combination, phrasal verb** and **separable verb** as possible terms. Each line of theoretical linguistics has a preferred reference term for this structure, which is the focus of Studies 3 and 4, as well as its own definition of this structure's properties (see Alejo González, 2010 for an overview).

In Cognitive Linguistics the term **multi-word verb** refers to any combination of a verb with an adverbial particle or a preposition that functions as a single unit. The meaning of a multi-word verb cannot be deduced from its parts. In the strict definition by Quirk, Greenbaum, Leech, & Svartvik (1985), it can further be subdivided into **phrasal verbs, prepositional verbs** and **phrasal prepositional verbs.** Unlike in generativism, particles and prepositions are not differentiated in cognitive approaches. Particles are defined as linguistic elements whose core meaning can be derived from their prototypical use as spatio-temporal adverbs and that can receive additional meaning through metaphorical use or experiential correlation.

The term **verb particle construction,** used in Generative Linguistics and Construction Grammar, distinguishes particles from prepositions and this distinction is achieved through a variety of syntactic tests. Constructionist approaches additionally do not distinguish between particles and freely combining adverbs, but particles are considered to have a more idiomatic meaning than freely combining adverbs. Ramchand & Svenonius (2002) assign the same syntactic properties to compositional (103) and idiomatic verb particle constructions (104).

(103a) They marched off the hangover.
(103b) They marched the hangover off.

(104a) They let up the pressure.
(104b) They let the pressure up.

As seen in the above examples, particles are homonyms of prepositions (and in German also of prefixes) and some linguistic traditions treat them as identical, while others make a clear difference. Dehé, Jackendoff, McIntyre, & Urban, (2002:3) suggest the following theory-neutral definition of Germanic particles: "(...) an accented element which is formally (and, often, semantically) related to a preposition, which does not assign case to a complement and which displays various syntactic and semantic symptoms of what may informally be called a close relationship with a verb, but without displaying the phonological unity with it typical of affixes." I will adopt this broad definition of particles in my thesis.

Regarding the syntactic structure of verb-particle constructions, two main approaches can be differentiated. The **complex predicate approach** to verb particle constructions takes (104a) as the paradigmatic one in which verb and particle are adjacent (Stiebels & Wunderlich, 1994; Neeleman, 2002). It assumes that verb and particle form a constituent and can be separated through syntactic processes. The noun always functions as the direct object of the verb (see the schematic in Table 6.1). This approach can be further subdivided regarding when verb and particle are combined. In the morphological version, this combination happens presyntactically in morphology or the lexicon and verb and particle form a complex head [$_V$ VPart]. In the syntactic version, verb and particle enter syntax as separate heads and are then combined into a phrasal constituent V' that excludes the object. The meaning of the verb particle construction is considered to be idiosyncratic and stored in the lexicon. To account for the productivity of verb particle constructions, the complex predicate analysis assumes the construction of complex verbs in syntax or the lexicon. In contrast, **small clause accounts** are based on sentences such as (103). They assume no selectional relation between verb and noun, capturing productive and compositional verb particle constructions equally well. Instead of being considered a direct object of the verb, the noun forms a small clause with the particle (see the right column in Table 6.1). Constructions like (104) would be treated as an idiomatic expression. The basic order of verb and particle is not agreed upon in the small clause account. Some assume the order in which verb and particle are adjacent as basic, some consider the noun-particle order the basic order. In both views, the non-basic order is derived through movement. The main difference between the complex predicate account and the small clause account is therefore whether the particle is incorporated syntactically with the verb or the noun. The two accounts were based on data from Germanic languages, especially English, Dutch and German.

	Complex predicate (Neeleman, 2002)	**Small clause (den Dikken, 1995)**
Base order:	$[_{VP} [_{V} \text{V Prt}] \text{DP}]$	$[\text{V} [_{SC} [\text{Prt DP}]]]$
Shifted order:	$[_{V'} \text{V} [_{VP} \text{DP} [_{V} t_{V} \text{Prt P}]]]$	$[\text{V} [_{SC} \text{DP}_i [\text{Prt } t_i]]]$

Table 6.1 Schematic representation of the complex predicate approach and the small clause approach to particle verbs

Combining a verb with a particle often activates a new or special meaning of the two parts. Particles are often ambiguous and adopt particular meanings only with certain verb groups, e.g. the particle *up* can express an approach with motion verbs (move up, run up), a decrease in size (roll up, fold up) or a completion (clean up, eat up). Certain meanings of a particular verb are only licensed when it is combined with a particle. The combination of the verb *say* and the particle/preposition *up* has no lexicalized meaning in English, but the same combination in German (*aufsagen*) expresses the meaning of *recite* and in Norwegian (*si opp*) it translates to *terminate* (as in *terminate a contract*). In the German example, the connection to the original meanings of verb and particle is still recognizable, whereas this relation is less transparent in the Norwegian example. The German verb-particle construction *abschreiben*, a combination of the verb *schreiben* ‘write’ and the particle *ab* ‘off’, has two lexicalized meanings, one can be translated as *write off/amortize*, the other as *copy*. In English, only one of the German meanings is rendered by a verb-particle construction, the other is a simple verb. The same is true for Norwegian, despite being made up of the same components, the first meaning (‘write off’) is rendered by the prefixed verb *avskrive*, while the second meaning (‘copy’) is expressed with a verb-particle construction *skrive av*. This cross-linguistic variation is an additional challenge for L2 learners as the same concept using the same words can either be a prefixed verb, a verb-particle construction or a simple verb.

If and how the more specific meaning of verb-particle constructions can be deduced from the individual parts of the construction is one of the most controversial questions in this field of research and has an impact on what is actually considered a proper exemplar of a verb-particle construction or a phrasal verb. One main issue regards the decomposability of these verb-particle constructions and how they relate to idioms. Generativism traditionally views these structures as linguistic anomalies and places them close to idioms as their meaning cannot be established in a bottom-up process. Instances in which the meaning of a verb-particle structure can be established from its parts, e.g. *go up,* are strictly not verb-particle constructions in this view. The opposite view held by Cognitive Linguistics is that all phrasal verbs can be placed on a continuum of decomposability and keep some of the meaning of their parts. The different meanings of a particle in Cognitive Linguistics are connected in a semantic network with the basic meaning at its core and the different extensions in meaning derived from it. In general, it can be said that some verb + particle combinations are more idiomatic and have a more opaque meaning that cannot be deduced from its parts, (e.g. *carry out, si opp, aufhören* ‘cease’, literally: ‘hear up’), while others show a productive pattern and their meaning can be derived more or less directly from their components (e.g. *go up, abschreiben*). This debate surrounding decomposability and the derivation of meaning relates directly to the questions of processing and storage of verb-particle constructions. Generativism assumes that opaque, idiomatic structures are stored in long-term memory and are retrieved as wholes, while transparent structures are built online. Usage-based approaches consider frequency as an additional factor. They posit that frequently used phrasal verbs are stored in long-term memory, even if they are fully decomposable, a view that is compatible with recent neurolinguistic evidence (Cappelle, Shtyrov, & Pulvermüller, 2010).

The Germanic languages investigated in this thesis show some variation with regard to the placement of the particle and the overall syntactic form of the entire construction. This variation is partly determined by the VO/OV parameter. In VO orders the particle is always separated from the verb, i.e. in a postverbal position. This applies to all tenses and moods in Norwegian and to V2 contexts in German (see Table 6.2 below). While the VO order causes the particle to be in a postverbal position, it does not determine the order of particle and object. Norwegian has a flexible order allowing the particle to either precede or follow the object. The German order on the other hand is fairly rigid in that the particle mostly follows the object. In OV orders the particle is always preverbal which applies to non-finite forms, verb clusters and embedded sentences in German and is never found in Norwegian.

	Norwegian		German	
Present	VO	Jon **drikker opp** vinen.		
		Jon **drikker** vinen **opp**.	VO	Johann **trinkt** den Wein **aus**.
Preterite	VO	Jon **drakk opp** vinen/vinen **opp**.	VO	Johann **trank** den Wein **aus**.
Perfect	VO	Jon har **drukket opp** vinen/vinen **opp**.	OV	Johann hat den Wein **ausgetrunken.**
Modal	VO	Jon må **drikke opp** vinen/vinen opp.	OV	Johann muss den Wein **austrinken.**
Embedded	VO	...at Jon **drikker opp** vinen.	OV	...dass Johann den Wein **austrinkt**.

Table 6.2 Overview of particle verbs in Norwegian and German, example: “John drinks up the wine.”

As I need a term that fits Norwegian and German verbs equally well, I will use the term *particle verb* when referring to verbs that are combined with a particle. This particle can be either attached to the main verb as in the German infinitives or it can follow the main verb as in Norwegian. My use of the term *particle verb* covers expressions with either a more composable or more opaque meaning as decomposability was not a factor that was manipulated in the design of the following experiments.

6.2 Literature review

6.2.1 Studies on object shift and particle shift

The syntactic background of object shift and particle shift will be discussed more in depth in section 7.1., but for the sake of order, experimental evidence will already be reviewed at this point. For now, it is enough to know that object shift concerns the movement of pronominal objects (*den*, 'it') across negation (ikke, 'not') or sentential adverbs (altid, 'always'), e.g. *Jon spiser den ikke* 'John does not eat it' or *Jon spiser den altid* 'John always eats it'. A non-pronominal object would always occur after the negation or the sentential adverb. Particle shift concerns the movement of objects (*his coat*) across the particle in particle verbs, e.g. *John takes off his coat* (no shift) vs. *John takes his coat off* (particle shift). Most studies on these phenomena are acceptability judgments, often of the kind criticized in section 2.4.2 as they are often only provided by the author and other colleagues.

6.2.1.1 Particle shift and particle verbs

Svenonius (1996a) compared judgment data on particle shift constructions in English, Norwegian and Icelandic. The judgments seem to have been collected in a rather informal manner and only very little information on the raters is provided. I will only report the Norwegian data here. Svenonius found a general preference for the particle-object (Prt-DP) order in some speakers that could be overridden by intonation or particle modification. The Prt-DP order was also preferred in contextless sentences with novel, indefinite objects: (105a) was preferred over (105b).

(105a) Jeg blaser opp ballonger.
I blow$_{\text{PRES}}$ up balloons$_{\text{INDEF}}$
'I blow up balloons.'

(105b) Jeg blaser ballonger opp.
'I blow balloons up.'

The same preference was also found when the particle verb had been introduced in the context and for impersonal constructions like (106).

(106) Det ramlet ned tre flasker.
there tumble$_{\text{PRET}}$ down three bottles$_{\text{INDEF}}$
'Three bottles tumbled down.'

In sentences with a given object, both orders are equally acceptable with a small preference for the DP-Prt order. A strong preference for the DP-Prt order emerges for epithetic objects as in (107).

(107) Q: How will Ingrid and Turid get here?
Vi skall plukke jentene opp.
we shall$_{\text{PRES}}$ pick girls$_{\text{DEF}}$ up.
'We will pick the girls up.'

In all contexts presented by Svenonius, both orders are equally grammatical from a syntactic point of view and the preferences never ruled out the other option. He concludes that the choice of one order over the other is a stylistic and extragrammatical one based on factors such as information structure or sentential intonation.

Production data from L2 speakers is available for Swedish-learning German native speakers (Bohnacker, 2006a, 2007). Bohnacker followed German L2 speakers of Swedish who had been living in Sweden for a number of years (between 3 and 15 years) and were exposed to the Swedish language on a daily basis through their work. Particle placement was one of many syntactic constructions investigated in this study that recorded naturalistic language production in a longitudinal design with up to three measurements. In Swedish, the particle always precedes the object regardless of object type (108a). In German, the particle usually follows the object and is often found in a sentence-final position (108b). The examples below are taken from Bohnacker (2007:36), emphasis added by me.

(108a) Nu tar jag **bort** diagrammet. (Swedish)
now take I away chart-the

(108b) Jetzt nehme ich das Diagramm **weg**. (German)
now take I the chart away
'I'll take the chart off now.'

Bohnacker found a prolonged non-native production of sentence-final particles in the L2 speakers. After 3 years of residence in Sweden, the non-targetlike order was predominant in all participants in finite and nonfinite constructions ranging from 57% to 100% of all production of particle verbs. Nonfinite verb placement was at 99.5% target-like production at the same point. Revisiting the majority of these participants again after another 3 years, this percentage had dropped for all participants and was now at 23-82%. Four participants were recorded at even later points, at 9 and 15 years of residence in Sweden and three of these participants produced non-targetlike particle verb orders in less than 10% of the cases. Overall, there was a developmental trend towards more native-like behavior for each participant, but the timeline stretches over nearly a decade of everyday exposure. Participants started out predominantly using the ungrammatical order that is similar to the order in their native language. This phase was followed by a stage in which both orders co-exist and the factors determining the choice of either order are unclear. Finiteness and particle type did not play a role as the same particle was used in target and non-targetlike constructions in finite and nonfinite contexts. Bohnacker argues that the prolonged non-targetlike production of DP-particle orders is due to full transfer from the L1 grammar, since the use of L1 German grammar regulates the sentence-final position of particles in V2 contexts. Bohnacker also suggests that it takes L2 learners with a German background nearly a decade to adjust their L2 grammar appropriately. This is a lot more immersive exposure than many participants in other L2 studies have. The transfer explanation is slightly problematic as there is no second L2 group with a differently structured native grammar that could show a different outcome of transfer. If the German L2 group fully transferred their grammar with regard to particle verbs, they should also produce unsplit verbs in non-V2 contexts which is apparently not the case as there was no effect of finiteness reported by Bohnacker. Instead of the full transfer suggested by Bohnacker, it seems more like a selective transfer (if transfer at all) that only applies to the position of the already split particle independent of the syntactic context, although the German grammar regarding particle verbs clearly specifies that splitting cannot be separated from the V2 context. As this was only a production study, no conclusions can be made whether the processing of these particle verbs develops in a similar way.

6.2.1.2 More general object shift

The study on various types of object shift in Swedish by Josefsson (2010) was similar to the Svenonius study above with regard to methodology. It provides more information about the raters, but the majority are nevertheless trained linguists. Another methodological concern is that the 20 items were all presented in a within-subject design, i.e. each subject saw all possible permutations of the same sentence, which may have influenced the results. The small amount of items with five data points per condition per participant might not have provided enough statistical power for more reliable results. Josefsson investigated short pronominal object shift in which an object moves across a negation or a sentential adverb, and long pronominal object shift in which the weak object pronoun *mig* appears between the verb *lärde* and the subject *Maria*, as in example (109a) below. Example (109b) represents the unshifted order.

(109a) Det här lärde mig Maria igår.
this here taught me Maria yesterday
'Maria taught me this yesterday.

(109b) Maria lärde mig det här igår.

For monotransitive sentences, Josefsson found a preference for the shifted pronoun over the unshifted pronoun when the pronoun was monosyllabic (*den* 'it'). For the disyllabic pronoun *honom* 'him' there was no preference for either order. No ordering preference was found when a disyllabic pronoun occurred with two sentence adverbials. For ditransitive verbs, Josefsson found again no statistically robust difference between the orders involving two pronominal objects and one sentence adverbial, as long as the indirect object preceded the direct object. A numerical trend suggested that the order indirect object > sentence adverbial > direct object was preferred. Any order in which the direct object preceded the indirect object was ruled ungrammatical. The results on long object shift suggested that it depends on the verb involved as there were opposite preferences for different verbs. Josefsson concludes that object shift is optional in Swedish and that several factors such as verb type, information structure, prosody and personal preferences play a role. The latter fact is reflected in the great variability between the ratings of the 26 participants.

One exception to the general trend of studying this topic through grammaticality judgments is an ERP study on object shift in Swedish by Roll, Horne, & Lindgren (2007). It investigated the influence of semantic and prosodic prominence on the processing of ungrammatical object shift comparing definite and indefinite full NPs. As pronouns easily shift across negation, it has been proposed that object shift applies mainly to easily accessible and unstressed objects. While definite and indefinite full NPs do not shift in Swedish, the authors argued that definite full NPs in a shifted position should be easier to process than indefinite full NPs as the latter always refer to new, relatively inaccessible information and receive stress. The behavioral data from an acceptability judgment task showed no difference between the two types of full NPs as participants accurately judged them as ungrammatical at more than 98%. There was a difference in the ERP data, though, indicating that indefinite full NPs were indeed harder to integrate than definite full NPs.

6.2.2 Studies on non-local dependencies

Particle verbs constitute a type of non-local dependency that has received some scientific attention in recent years, both from a purely theoretical point of view as well as through experimental approaches. Much of the experimental evidence in processing is mainly centered around the mental representation of particle verbs (e.g. Cappelle, Shtyrov, & Pulvermüller, 2010) or the inference of their meaning (e.g. Blais & Gonnerman, 2013). Participants are often native speakers, though the Blais & Gonnerman study is an exception. L2 studies are often production studies (see Bohnacker, 2006, 2007 in section 6.2.1) and a common finding seems to be that L2 speakers tend to avoid particle verbs (Dagut & Laufer, 1985; Liao & Fukuya, 2004; Siyanova & Schmitt, 2007). Authors have claimed that particle verbs are problematic in L2 acquisition due to their variability between transparent and opaque meanings and the fact that they can often be replaced with a simple verb without massive changes to verb meaning. The studies that will be reviewed in this section focus mainly on ordering preferences and their consequences for processing. Non-local dependencies in general have been found to cause processing difficulties that are reflected in slower online reading times and lower offline performance. Memory-based processing accounts explain this by positing that an increase in the distance between the two elements of the dependency makes retrieval from memory more challenging. As the number of processing studies investigating particle verbs is small, I will also review experiments on related non-local dependencies.

6.2.2.1 Particle verbs and non-local dependencies in English

In order to account for differences in the production of word orders with the same truth-value content, such as particle verbs, or extraposed and adjacent relative clauses, Hawkins (1994) proposed the **Early Immediate Constituent (EIC) principle** that was later updated in his **Minimize Domains** (MiD, Hawkins, 2004) theory. The EIC/MiD is a principle that 'assigns a quantified and gradient preference to one structure over another' (Hawkins, 2011:249). It assesses the overall complexity of structures that leads to their selection in production. As production is at the center of EIC/MiD, its main interests are principles that allow high speed in communication while at the same time minimizing the processing effort. Hawkins proposes four principles that are involved in this task:

(1)	Early Immediate Constituents	Linear orders that minimize Phrasal Combination Domains (PCDs) by maximizing IC-to-word ratios are preferred
(2)	Maximize On-line Processing	Linguistic forms are selected and arranged in order to allow the earliest access possible to the highest amount of syntactic and semantic structure
(3)	Minimize Domains	Connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties should be minimized
(4)	Minimize Forms	Formal complexity of linguistic forms and the number of forms with unique conventionalized property assignments should be minimized

Table 6.3 Four principles of the Minimize Domains theory

In Hawkins' theory, Phrasal Combination Domains (PCDs) are used to evaluate competing orders against each other. PCDs contain a mother node, for example a VP, and all the immediate constituents (ICs) of this mother node. The amount of material that has to be processed to construct the mother node and assess the number of its immediate constituents varies between competing orders. Less material results in faster and more efficient phrase-structure processing. IC-to-word ratios are used as a numerical measure to compare orders and are exemplified below (example taken from Lohse, Hawkins, & Wasow, 2004):

(110a) Joe $_{VP}$[looked up $_{NP}$[the number of the ticket]]
1 2 3
VP PCD: IC-to-word ratio of 3/3 = 100%

(110b) Joe $_{VP}$[looked $_{NP}$[the number of the ticket] up]
1 2 3 4 5 6 7
VP PCD: IC-to-word ratio of 3/7 = 43%

In the case of a transitive particle verb, there are three immediate constituents (ICs): the verb, the particle and the first word of the object DP. In example (110a), three words are needed to encounter all three ICs, resulting in a perfect IC-to-word ratio of 100%. In example (110b) on the other hand, the entire NP needs to be processed to reach the particle and with it identify the last IC, reducing the IC-to-word ratio to 43%. The adjacent order of example (110a) should therefore be preferred over the non-adjacent order of example (110b) as it allows an earlier recognition of the phrase-structure and there is less additional processing alongside phrase-structure processing. Hawkins derives the following order predictions from his theory:

Adjacency is preferred when…	… distance between the two elements of a discontinuous dependency **increases** … weight and complexity of the second element of the dependency **decreases**
Non-adjacency is preferred when…	… distance between the two elements of a dependency **decreases** … weight and complexity of the second element of the dependency **increases**

Table 6.4 Adjacency predictions derived from the EIC

The EIC has not only been used to examine variations of adjacency, but also to account for word order variations of ditransitive constructions. Hawkins makes an additional prediction, namely that verbal position interacts with length. Verb-first and verb-second sentences should show a strong preference for the “short before long” principle as the recognition of the constituents would otherwise be unnecessarily delayed. Verb-final sentences, like German subordinate clauses, should show a weaker preference for the “short before long” principle as the recognition of all constituents has to be delayed until the final word of the sentence in any case.

A corpus study on English particle verbs by Lohse, Hawkins, & Wasow (2004) investigated the occurrence of adjacent and non-adjacent orders of these verbs against the background of the EIC/MiD. They found that selection of either order was sensitive to differences in efficiency between two structures. They also found evidence for subtle locality effects in syntactic and semantic processing domains. The ratio of non-adjacent orders of verb and particle showed a first steep drop at an NP length of three words and a further one to 3% at an NP length of five or more words. The authors did not find a single instance of a non-adjacent order for NPs longer than eight words. Lohse et al. proposed a classification of particle verbs according to the dependency of the particle or the verb leading to four different groupings. At the one extreme are combinations in which both verb and particle are processed independently resulting in a fully compositional meaning. At the other extreme are combinations in which verb and particle depend on each other, rendering the meaning opaque. In between are two combinations in which one member is dependent, while the other is independent. Similarly to the VP PCD explained above, dependent particles have a lexical dependency domain (LDD) that should be kept as short as possible predicting a stronger preference for adjacent orders than in independent particles. The authors did find a significantly lower non-adjacent ratio for dependent particles than for independent particles (16% vs. 42%) that was independent of NP length. The dependency status of the verb does not affect ordering preferences as dependent and independent verbs both need to access the direct object, while only dependent particles need to access the verb. Lohse et al. found this prediction to be supported as there was no difference in non-adjacency ratio between dependent and independent verbs. Some restrictions to this finding apply when the noun is modified. The adjacent order is preferred whenever the head noun is modified postnominally, while the non-adjacent order is preferred when the object NP is modified prenominally.

Processing studies of non-local dependencies often aim to increase the processing load by increasing the distance between the dependents, inserting material that is similar to the word that needs to be retrieved or by adding a second task. Fedorenko, Woodbury, & Gibson (2013) wanted to investigate the possibility of a decrease in processing load by facilitating the retrieval of the dependent. This facilitation is also predicted by memory-based processing accounts. The experiment involved a dual-task paradigm, combining self-paced reading with a memory task. The structure investigated were cleft sentences in which either the subject (local dependency) or the object (non-local dependency) had been extracted.

Subject-extracted condition, local dependency
(111) It was John who consulted Ellen in the library.

Object-extracted condition, non-local dependency
(112) It was Ellen who John consulted in the library.

In the additional memory task, participants had to remember either the subject, the object or an unrelated control word. In a fourth control condition, participants did not complete a memory task. In the object-memory-word object-extracted conditions, the object's salience is improved through the memory task and retrieval was predicted to be sped up compared to the no-memory-word control condition. The subject-memory-word conditions tested an alternative hypothesis that the encoding of a word as topic leaves a stronger memory trace that is further facilitated by the memory task predicting faster reading times in the subject-memory-word subject-extracted conditions compared to the no-memory-word control condition. The control-memory-word conditions served as a baseline for the general effect of the memory task on the reading times. The object-memory-word object-extracted condition showed faster reading times and higher comprehension rates compared to the other three object-extracted conditions. In the reading times, the extraction advantage of subject-extracted sentences was even eliminated. A similar effect of subject-memory-word on the reading times of subject-extracted sentences seemed to support an encoding-based hypothesis that a general match between memory word and sentence topic would facilitate retrieval. However, an absence of this facilitation in the comprehension data supports a retrieval-based hypothesis. Increasing the distance between the dependents, or secondary tasks taxing the retrieval of the first dependent, are known to hinder retrieval, whereas enriched representations of the dependent and a more prominent syntactic position facilitate retrieval from memory.

6.2.2.2 Particle verbs and non-local dependencies in German and Dutch

Experimental research on word order in German involving particle verbs dates back to the 1970s when Köpcke & Zubin (1979) investigated the perception of sentences in which certain parts of speech were placed on either side of the particle (see examples (113a) and (113b) below adapted from Köpcke & Zubin, 1979). In example (113a) the prepositional phrase (*auf den linken Flügel* 'to the left wing') follows the particle (*ab*) and particle and main verb (*spielen*) are adjacent. In example (113b), the prepositional phrase intervenes between main verb and particle. The particle ends up in verb-final position and is no longer adjacent to the main verb creating a discontinuous dependency. In (113a), the prepositional phrase has been placed outside the sentential bracket (*ausgeklammert*), while it has been placed inside (*eingeklammert*) in (113b). In the examples, the asterisk does not indicate ungrammaticality, but rather the location of a click sound played in the sentence.

(113a) Der Mittelstürmer spielt ab * auf den linken Flügel.
the center forward plays off to the left wing
'The center forward passes to the left wing.'

(113b) Der Mittelstürmer spielt * auf den linken Flügel ab.

Using a click paradigm, Köpcke & Zubin (1979) investigated the perception of these two orders by German native speakers. Click experiments assume that processing capacities during sentence perception are limited and that speakers will try to alleviate processing demands, if possible. Participants have to complete two tasks simultaneously: hearing/processing of a sentence and monitoring of a click sound. After hearing sentence and click, participants have to mark where in the sentence the click sound was heard. Due to the dual nature of the task, the pressure on the processing capacities causes the participants to misperceive the location of the sound and place it along the structural segmentations of the input. For instance, a click played during the reading of an NP will afterwards likely be placed before the whole NP. In this study, sentences were presented either in isolation or with a context. The critical structural border was assumed before the adverbial phrase in both placement conditions. Clicks were placed either at the border (central), three positions to left, or three positions to the right, resulting in three click conditions. To account for all possible click positions in a sentence, syllables and spaces between words were counted resulting in 22 possible positions for examples (113a) and (113b) above. In order to make sure that participants processed both the content of the sentence and the position of the click, participants had to answer content questions in addition to the click localization task. The results showed that the click position was misremembered in nearly 80% of all analyzed cases, with a tendency to place the click farther to the left. In the externalized condition, there was a border effect before the adverbial phrase in the central and the right click condition. This effect was absent in the internalized condition. The highest accuracy for click perception was reached for the central condition in which the click was placed on the proposed border. The authors suggest a similar role for particles in externalized sentences and function words in the perceptional process, evoking an internal perceptional border.

Uszkoreit et al. (1998) and Konieczny (2000) used several different methods to investigate the predictions made by the EIC for German extraposed relative clauses as in (114a) and (114b) (from Uszkoreit et al., 1998).

(114a) Er hat [$_{VP}$ das Buch, das er gestern erst gekauft hat, heute gelesen.]
‘He read the book that he only bought yesterday today.’

(114b) Er hat [$_{VP}$ das Buch heute gelesen, das er gestern erst gekauft hat.]
‘He read the book today that he only bought yesterday.’

The results of the three methods – corpus analysis, acceptability judgments and self-paced reading – converge only partly. The corpus data showed a sharp decrease of the extraposed order, if the intervening material exceeded more than three words, while it was the predominant order if the distance was only one or two words. The judgment data showed a strong general preference for the adjacent order. Acceptability for the extraposed order decreased with increasing distance, while acceptability for the adjacent order decreased with increasing weight and complexity of the relative clause. The self-paced reading data showed generally faster reading times for the adjacent order and a main effect of distance for both orders, but no effects of relative clause weight or complexity. While the results of the corpus study matched the predictions of the EIC, the self-paced reading data did not show an effect of clause weight, and the acceptability data showed neither an effect of weight nor of distance. Similar differences were found in a study on English extraposed relative clauses by Francis (2010). Hawkins (2011) explains these divergences from the predictions made by the EIC by the fact that the three measures investigate different aspects of language processing and use. Overall complexity as measured by the EIC is not necessarily reflected by the processing ease at individual points of a sentence and therefore its predictions do not seek to explain the self-paced reading data. As for the acceptability data, Hawkins (2011) suggests the influence of a normative bias that introduces a strong preference for the adjacent order.

In order to investigate the processing of prefixed verbs in relation to their matrix verbs, Smolka, Komlósi, & Rösler (2009) used an overt visual priming paradigm in which verbs were either primed by a semantically related verb, a morphologically and semantically related verb, a solely morphologically related verb, orthographically related or an unrelated verb (see Table 6.5 below).

Target	**Prime**	**Type**
kommen 'come'	*nahen* 'approach'	semantically related
	mitkommen 'come along'	morphologically and semantically related
	umkommen 'perish'	morphologically related
	kämmen 'comb'	orthographically related
	schaden 'harm'	unrelated

Table 6.5 Prime-target combinations from Smolka, Komlósi & Rösler (2009)

The authors report two overt priming experiments in which the prime was displayed for 300ms and 1000ms respectively. They found morphological priming that was not related to semantic transparency in both experiments, i.e. *mitkommen* and *umkommen* had the same facilitative effect for the target *kommen*. Semantic priming was found, but the effect was about half the size of the morphological effect. Orthographically related primes resulted in significant inhibition in the short display time (300ms) and no significant inhibition in the long display time (1000ms) suggesting a decline of inhibitory influence over time. The results of this study suggest that German native speakers associate prefixed verbs with the bare matrix verb independent of the semantic relation between the verbs.

Particle verbs often occur in forms in which the particle and the verb are separated from each other by constituents of any kind of length, but the verb itself often already carries meaning in the absence of the particle. The parser has to keep the verb in memory in order to be able to relate it to an eventually following particle. Piai, Meyer, Schreuder, & Bastiaansen (2013) investigated the role of working memory in the processing of Dutch particle verbs using ERPs. Their experiment was based on two assumptions. First, the processing of particles involves primary and secondary lexical access. The verb needs to be processed to assign its syntactic and semantic properties to the particle and the full interpretation of the verb can only take place once the particle has been encountered. Second, as the bare matrix verbs can also appear without the particle, the dependency is uncertain, because it is unclear whether a particle will follow unlike in other syntactic dependencies. The authors investigated the role of working memory in the dependency formation and that of long-term memory in lexical access. One experimental factor was the number of different particles a verb could take: no particle at all, two particles or more than five. If the number of possible particles plays a role in the prediction of an upcoming verb particle dependency, there should be a difference between the verb with two particles and those with five. If only the mere possibility of a following particle is relevant, the verbs with particles should not differ from each other. The second experimental factor manipulated was whether the particle and the verb formed an existing particle verb. One condition combined verbs that allow a particle with a fitting particle creating an existing particle verb (well-formed condition). Another combined verbs with a particle that they can take, but the resulting particle verb does not fit the context (semantic violation condition) or with a particle creating a non-existing particle verb (morpholexical violation condition). Finally, the morpholexical violation condition combined a verb with an illicit particle to create a non-existing particle verb. If there is a secondary lexical access at the particle, the authors expected an N400 in the morpholexical violation condition reflecting a more effortful lexical access due to the ungrammaticality.

verb= *bellen* 'call' (combines with two particles)		
well-formed condition	*af* 'off'	Wij bellen de afspraak van vanmiddag af. 'We call the appointment of this afternoon off'.
semantic violation condition	*terug* 'back'	Wij bellen de afspraak van vanmiddag terug. 'We call the appointment of this afternoon back'.
morpholexical violation condition	*toe* 'to'	Wij bellen de afspraak van vanmiddag toe. 'We call the appointment of this afternoon to'.

Table 6.6 Experimental conditions used by Piai et al. (2013)

The authors found no difference in ERP signature between verbs taking two or five particles, but a general difference between either of the groups of particle verbs and verbs accepting no particles at all. They suggest that L1 sentence processing is sensitive to the possibility of a particle dependency later on in the sentence and that this sensitivity is not modulated by the number of possible particles associated with a verb. They found a left anterior negativity for the processing of particle verbs which they interpreted as reflecting the costs for storing the verb in working memory. The authors found a graded N400 effect for the manipulation of what kind of particle verb was presented. Both violation conditions produced a larger N400 effect than the well-formed condition, and the largest effect was found in the morpholexical violation condition. The graded N400 was interpreted to reflect a mismatch effect in the semantic violation condition as the particle verb could be found in the mental lexicon, but could not be integrated into the context. In the morpholexical violation condition, it was interpreted to reflect the unsuccessful attempt to access a non-existing entry.

6.3 Summary

Despite not necessarily involving object movement or non-canonical word orders, the study of particle verbs is nonetheless informative of native and non-native syntactic processing as their surface form is influenced by the OV/VO parameter and can vary depending on the syntactic context. According to some theoretical approaches, similar ordering constraints apply to the order of particle and object as to the two objects in ditransitive sentences that were the subject of Study 2. The existing experimental evidence comes mainly from the area of acceptability judgments, corpus studies and production data, usually with native speaker participants. Studies using online methods such as priming and EEG have found that native speakers automatically associate prefixed verbs and particle verbs with the matrix verb independently of a semantic connection between verb and particle. This suggests that the degree of a particle verb's opaqueness is not highly influential on native processing. Native speakers appear also to store verbs during online processing if a particle could possibly follow later on in the sentence. This effort is not related to the number of possible particles that the verb can take. For non-native speakers, it is mainly known that they tend to avoid particle verbs in production and struggle with the correct ordering of verb, particle and object. Online, non-native processing of particle verbs is comparatively unstudied as processing studies often focus on the interpretation of figurative particle verbs (e.g. Paulmann, Ghareeb-ali, & Felser, n.d.). The two studies presented in the following chapters investigated to what extent non-native speakers of Norwegian have the same ordering preferences for particles and objects as native speakers (Study 3), and whether non-native speakers of German can integrate the rules for splitting particle verbs into their online processing (Study 4).

7 Study 3: Particle placement in Norwegian

In this experiment I investigated the role object length plays on the processing and acceptability of shifted objects in constructions containing particle verbs. As already mentioned in Section 6.1, Norwegian particle verbs allow the object to be placed either before the particle or after, and both orders are claimed to be equally acceptable. According to the Norwegian Reference Grammar (Faarlund et al., 1997) object length plays a major role in developing a preference for the object-final order. So far, most publications on object shift and particle shift in the Scandinavian languages deal with the theoretical side of this phenomenon along with some production studies in L1 and L2 acquisition. All judgments regarding acceptability and/or grammaticality were based either on the author's native speaker intuitions or judgments by small groups of other native speakers, often also trained linguists (see literature review Section 6.2.1).

The acceptability judgments in Experiment 3a were collected from a larger group of Norwegian native speakers, most of which were linguistically naïve.[13] The judgments were also collected in a more systematic manner using a Latin Square design and a systematic manipulation of the experimental variable object length. The aim of this experiment was to evaluate conclusions based on previous, less systematic offline judgments and to investigate whether the statements of descriptive grammar can be supported by data from an online processing study. In addition to the evaluation of claims regarding native speaker processing, this study will investigate whether non-native speakers adhere to the same ordering preferences in acceptability judgments and online processing as native speakers.

Theories like the EIC predict that the particle > object order is preferred as it allows the earlier assessment of all important constituents and unambiguous interpretation of the particle verb, thereby facilitating processing. Alleviating the processing resources is also in the best interest of the L2 parser, especially if we assume that L2 speakers need to allocate more resources to parsing than L1 speakers. As the EIC is based on general cognitive mechanisms, it does not make

[13] Due to limitations of available participants, some of the participants (n=13) were enrolled in subjects such as Scandinavian or English studies. It is not possible to estimate their linguistic knowledge of the construction investigated. However, none of the students were enrolled in the linguistics program.

different predictions for L1 and L2 speakers and to my knowledge, there has been no experimental investigation of the application of the EIC in L2 speakers. The ordering of particle and object is governed by similar principles as the ordering of the objects in ditransitive sentences and in study 2, L2 speakers were not particularly sensitive to the difference between the two orders, neither in the rating task nor in online processing. Study 3 aims to investigate whether L2 speakers are insensitive to order variations of this type and whether they show evidence of considering a possible particle verb interpretation in online processing when they encounter a verb that can be used with a particle as the Dutch L1 speakers did in the study by Piai et al. (2013). Preparing for a possible particle verb by storing the verb and constructing a filler-gap dependency at the particle should be visible in the self-paced reading data and could give insight into the mental representation of particle verbs in L2 speakers.

This chapter is structured as follows: First, I will give more detailed information on the theoretical background of object and particle shift in Norwegian (Section 7.1.). As the Norwegian Reference Grammar and previous works state a decline in acceptability of shifted heavy noun phrases all materials were pretested and evaluated by native speakers to avoid the presentation of shifts that are perceived as ungrammatical. This pilot study is reported in Section 7.2. Section 7.3 reports Experiment 3a – the acceptability rating task, and Section 7.4 reports the results of Experiment 3b the corresponding self-paced reading task. Section 7.5 provides a summary of the chapter and discusses the results of the experiments.

7.1 Background: Object and particle shift in Norwegian

Shifting the position of an object or subject – no matter if it is a full NP or a pronoun – within a sentence is not uncommon in Norwegian or any other of the Scandinavian languages. The shift of an object to the front of the sentence – object topicalization – was investigated in Study 1. Study 3 focuses on shifts that occur in the Norwegian midfield where subjects and objects can both be subjected to shifts that have different properties. Despite apparent similarities, these shifts are distinct from scrambling that was investigated in Study 2 (Anderssen, Bentzen, Rodina, & Westergaard, 2010). For example, scrambling allows multiple possible landing sites, whereas object shift has a fixed landing site and depends on the movement of the finite main verb (Vikner, 2006). Subject shift applies to pronominal and full NP subjects in non-subject initial sentences and has been interpreted to either reflect specificity or information structure. Subjects with a specific reading and given (often unstressed pronominal) subjects precede the negation or sentence adverbs in (115a) and (115b). Weak subjects with a non-specific reading and new or focused subjects (also pronouns with contrastive focus) follow negation in (116a) and (116b). The following examples are taken from Anderssen et al. (2010), with subject printed in bold, and italics in original.

(115a) Igår leste **Jon** ikke boka.
yesterday read$_{\text{PAST}}$ Jon not book-the
'Yesterday Jon didn't read the book.'

(115b) Igår leste **han** ikke boka.
Yesterday read$_{\text{PAST}}$ he not book-the
'Yesterday he didn't read the book.'

(116a) Igår leste ikke **Jon** boka.
yesterday read$_{\text{PAST}}$ not Jon book-the
'Yesterday *Jon* didn't read the book.'

(116b) Igår leste ikke **han** boka.
Yesterday read$_{\text{PAST}}$ not he book.the
'Yesterday *he* didn't read the book.'

Object shift occurs only in contexts of main verb movement (Holmberg's generalization, Holmberg, 1986) and is therefore not found in embedded clauses or main clauses with auxiliaries. It is also related to information structure and possibly the best-documented case of either shift type in Norwegian. It is linked to case assignment as only objects that have morphological case marking may be shifted. This means that object shift is ungrammatical for full NPs (117b) or non-specific indefinite object pronouns as they do not bear case marking. But it is obligatory for unstressed definite pronouns (117c) as pronouns in the unshifted position (117d) receive contrastive stress. In ditransitive constructions, direct objects cannot shift across indirect objects (see Vikner, 2006 or Broekhuis, 2008 for a more comprehensive overview)

(117a) Jon leser ikke **boken**.
Jon read$_{\text{PRES}}$ not book-the
'Jon does not read the book.'

(117b) *Jon leser **boken** ikke.

(117c) Jon leser **den** ikke.
'Jon does not read it.'

(117d) Jon leser ikke **DEN**.
'Jon does not read that one.'

In the literature, object shift has been compared to scrambling that was investigated in Chapter 5. Both involve the leftward movement of objects from inside VP to outside VP and the final landing site is determined by properties of the object. The difference between the two phenomena is the stronger restriction of object shift as scrambling can encompass several landing sites and extend beyond objects. Vikner (1994) for example explains the different syntactic properties of scrambling and object shift with different types of movement. He interprets scrambling in Western Germanic as A-bar-movement and object shift in Scandinavian as A-movement. With the exception of a few dialects object shift exists in all of the Scandinavian languages and its absence has been explained by the phonological properties (i.e. sentential intonation) of the dialects in question (Hosono, 2010a, 2010b). Despite the pan-Scandinavian presence of object shift, the extent to which pronominal objects are shifted is language-specific and shows some variation dependent on the pronoun's antecedent (see Andréasson, 2009 for a comparison of Swedish and Danish). The differences between Danish and Swedish that Andréasson found in her corpus study on pronominal object shift remind us that even for languages as closely-related as the Mainland Scandinavian languages, generalizations across languages are often not possible. Therefore, any studies on a language other than Norwegian – such as the studies reviewed here on Swedish and Danish – can only give a vague idea about the properties of Norwegian, and only serve as a crutch in the absence of data on Norwegian. This is especially important to remember because, despite a large amount of theoretical work on object shift, there are to my knowledge no experimental studies, and no processing studies, on Norwegian object shift. Filling this void is an important objective of this study.

Related to this general object shift is the phenomenon of 'particle shift' that describes the shift of a full NP or pronominal object across the particle of a particle verb. Particle shift of pronouns follows the same rules as general object shift or pronouns, i.e. unshifted pronouns receive contrastive stress and unstressed pronouns have to shift (see 118c-e below). Particle shift of full NPs has no corresponding analogy in general object shift and seems to be entirely optional. Unlike object shift proper, particle shift does not depend on previous verb movement and it differs across the Scandinavian languages. The shifted object is the only grammatical option in Danish (similar to 118b below), while Swedish only allows the unshifted option (similar to 118a below). In Norwegian and Icelandic on the other hand both positions are grammatical and particle movement is optional (Holmberg & Platzack, 2005). As all four of these languages are OV languages, the difference in the availability of particle shift is attributed to a difference in syntactic status of the particle. Hróarsdóttir (2008) analyses the particle as a phrase in Danish, as a head in Swedish and as alternating between phrase or head in Icelandic and Norwegian. A modified particle would always be considered a phrase in the latter two languages and cannot move.

(118a) Ole tar på **genseren**.
Ole put$_{\text{PRES}}$ on sweater-the.
'Ole puts on the sweater.'

(118b) Ole tar **genseren** på.
'Ole puts the sweater on.'

(118c) *Ole tar på **den**.
*Ole puts on it.

(118d) Ole tar den på.
'Ole puts it on.'

(118e) Ole tar på **DEN**.
'Ole puts that one on.'

Despite the intense theoretical discussion of object shift proper and also particle shift, there is no consensus over which of the two orders is the basic one and whether it is the particle or the object that is moved.[14] While theoretically possible, accounts that involve rightward movement of constituents are not pursued in the literature, leaving two possible basic orders and their derivations as shown in Table 7.1. Taraldsen (1983) takes the Prt-DP order as the basic order (118a) and derives the DP-Prt order through leftward movement of the object

[14] The possibility that both orders are base generated and there is no movement of either object or particle has not been suggested in the thirty years since the debate first reached a wider scientific audience.

(118b). In the basic Prt-DP order, particle constructions and ordinary PPs are constructed in a parallel fashion. Åfarli (1985) criticizes Taraldsen's approach as too limited as it can only be applied to particles that can actually be construed as prepositions. He does not treat particles as a separate word class, but rather analyses them as PPs, AdvPs or APs depending on the particle. As particle and object cannot be topicalized together, they also do not form a common constituent. He also does not consider the particle verb as base-generated from the lexicon and instead treats the particle as separate from the verb, which is problematic given the experimental evidence on the storage of particle verbs reviewed in the preceding section. In Åfarli's view, DP-Prt is basic and the Prt-DP order is derived through leftward movement of the particle, but he does not posit a trace for the moved particle. Åfarli interprets the derived word order as a grammaticalization of a causative reading as this is supposedly the only legitimate reading for the SVPredO order, while the basic SVOPred order can receive a causative and a non-causative interpretation. Åfarli also emphasizes the role of dialectal variation in the choice which PREDs actually behave like particles and that this choice seems to be governed by prosody (see similar approach to object shift proper by Hosono (2010a,b) above).

	Åfarli (1985)	**Taraldsen (1983)**
Basic	$[_{S}\ \mathrm{S}\ [_{VP}\ [_{V}\ \mathrm{V}]\ [_{NP}\ \mathrm{O}]\ [_{PP}\ [_{P}\ \mathrm{PRED}]]]]$	$[_{S}\ \mathrm{S}\ \mathrm{V}\ [_{PP}\ [_{P'}\ [_{P}\ \mathrm{Prt}]\ [_{NP}\ \mathrm{O}]]]]$
Derived	$[_{S}\ \mathrm{S}\ [_{VP}\ [_{V}\ \mathrm{V}]\ [_{XP=particle}\ \mathrm{PRED}]\ [_{NP}\ \mathrm{O}]]]$	$[_{S}\ \mathrm{S}\ \mathrm{V}\ [_{PP}\ [_{NP}\ \mathrm{O}]_{i}\ [_{P'}\ [_{P}\ \mathrm{Prt}]\ [_{NP}\ t]_{i}\ ]]]$

Table 7.1 Comparison of syntactic analyses of particle shift by Åfarli (1985) and Taraldsen (1983)

Both sentences in (118a) and (118b) have the same truth-conditional content and are supposed to be free variations of each other (Svenonius, 1996a). This apparent optionality in Norwegian does have some underlying rules though. One concerns the object type and its informational status similar to object shift proper. As seen in the examples (118c) and (118d) above, weak indirect object pronouns always precede the particle and also content adverbials. Focused or strong indirect object pronouns can follow the particle (118e). Full NPs, both focused and unfocused, can occupy both positions (118a) and (118b).

Not all particles are equally suitable for particle shift. Particles cannot be shifted if they are modified by an element such as *rett* 'right' or have complements. The latter type of particle would then be analyzed as prepositional (Svenonius, 1996a). Askedal (1982) differentiates between true verbal particles that occur with a particle verb and non-commutable prepositions (‘ikke-kommuterbare preposisjon’) that are part of transitive verb groups. These two groups belong to two different paradigmatic classes and have distinct semanto-syntactic properties: particle shift is only possible with true verbal particles. Norwegian verbs can occur with either a non-commutable preposition (e.g. *passe på* ‘look after’), a verbal particle (e.g. *drikke opp* ‘drink up’) or with both in a strict particle-before-preposition order (e.g. *se opp for* ‘look out for’), but never with two verbal particles or two non-commutable prepositions at a time. The particles themselves can either be analyzed as functional heads outside the corresponding small clause or as lexical heads of the small clause (Taraldsen, 2000).

Object length is often claimed as one of the main factors determining the availability of particle shift as ‘very heavy NPs’ cannot be shifted, but no statement is made specifying the actual meaning of ‘very heavy NP’ (Faarlund, Lie, & Vannebo, 1997:783; Svenonius, 1996a, 1996b). It is therefore not clear whether a ‘very heavy NP’ is just a very long object NP consisting of multiple words, for example several adjectives, or whether it has to be a complex phrase, e.g. an object NP followed by a relative clause. The decline in the availability of particle shift is explained by an increasing difficulty to parse the sentence-final particle and successfully establish the filler-gap dependency with the matrix verb, especially in cases when the matrix verb can also be used without the particle.

A length-based approach to constituent ordering like Hawkins' (1994) EIC principle that was introduced in earlier chapters is in line with the suggestion of the Norwegian Reference Grammar (Faarlund et al., 1997). Long fronted objects lead to an unfavorable constituent-to-word ratio as the parser needs to process the object in its entirety before reaching the particle and being able to compute the final number of constituents, whereas a fronted particle improves the constituent-to-word ratio as only the particle and the first word of the object need to be processed to posit the existence of the same two constituents. The EIC would also predict an almost linear decline in occurrence and acceptability of fronted objects with growing length compared to the non-fronted order. A quicker recognition and processing of the phrase structure relieves the processor of competing structures and is therefore less demanding for working memory.

In addition to length, several other variables that influence the order of particle and direct object have been found for English and were usually investigated in isolation. While interactions between these factors are highly likely, they are hard to accommodate in minimal pairs. Gries (2002) evaluated these variables with regard to their influence on processing ease and concludes that the particle > object order is felicitous for objects with high processing, while the order object > particle fits objects with low processing effort. Processing effort is determined by the variables listed in Table 7.2 below. An object causing high processing effort would be a long complex full NP that is stressed and contains new information. Gries assigns the biggest importance to morphosyntactic variables, followed by semantic and discourse-functional variables. Data from the British National Corpus also indicate an influence of modality of language use, as the particle > object order is more frequent in written data, while the object > particle order is more frequent in spoken data.

	particle > object	**object > particle**
Phonology	thoroughly processed, stressed referents	
Morphology :		
- Verb processing	immediate completion	
- NP type	lexical DOs	personal pronouns, referentially vague nouns
- Determiner	indefinite	definite
- Length + complexity	longer, more complex objects	shorter, less complex objects
Semantics:		
- Idiomaticity	idiomatic, opaque meaning	lexical meaning, stress on spatial meaning
- Animacy	no effect on processing ease	
Discourse-functional:	discourse-new or hearer-new information	given information
Other:		presence of directional adverbial
	high production and planning costs	low production and planning costs

Table 7.2 Variables influencing particle order, based on Gries (2002)

While there is no commonly agreed upon canonical order of particle and object in for Norwegian particle verbs, and the two orders are supposedly optional variations of each other, there are a number of factors that make one order more preferable than the other. The factor investigated in this study is object length. The object > particle order should be judged less favorably and should elicit higher reading times with increasing object length compared to the particle > object order.

Research questions:

Based on the theoretical background and the previous research on the processing of particle verbs and other discontinuous dependencies, the following research questions for study 3 will be addressed:

Q3.1 Do native and non-native speakers have a general preference for one order of particle and object over the other?

a) Are the two orders interchangeable variations of each other for short NP objects (but not necessarily for longer ones) as is suggested by the reference grammar?

b) Is the verb – particle – object order the preferred order, as proposed by the EIC and Gries (2002) which state that it is easier to process?

c) If present, is this preference visible both in the online self-paced reading data and the offline acceptability rating?

Q3.2 Does the object – particle order get less acceptable for longer objects as suggested by the NRG, the EIC and also Gries (2002)?

a) Is this decrease in acceptability visible in online processing tasks in which a longer object might tax processing resources more than a short object…

b) …and/or is the decrease visible in offline rating tasks?

Q3.3 Do L2 speakers of Norwegian with German as their native language show an advantage for the particle-final construction that is also present in their native language as was found in the production study by Bohnacker (2007)? Do they favor the object – particle order more than the Norwegian control group?

7.2 Pilot study

A pilot study was conducted to test for a possible cutoff point for the length of objects in shifted positions and to decide between the medium and long object for the final experiment to be included in the self-paced reading task.

Participants

The participants were 91 native speakers of Norwegian that had not participated in the pretest of study 1 reported in section 4.2. They were recruited either through personal contact or online and did not receive a reimbursement for their participation. The software used for this questionnaire recorded the participants' location based on IP address. Six participants had IP addresses outside of Norway (Denmark, Germany, Sweden, United States, UK), but this was not a criterion for exclusion. Data from six participants was excluded as they reported a second native language in addition to Norwegian (Danish, Swedish, Polish or English) and an additional two participants were excluded as they reported a diagnosed speech impairment.

The remaining 83 participants (male n=66) had an average age of 25.04 years (range: 15-56, SD: 6.56). Participants also provided information on their use of Norwegian. 77 participants (93%) reported Bokmål as their written variety of Norwegian. The participants spoke dialects from all over Norway: 28 spoke dialects belonging to the Western dialect group (34%), 30 spoke dialects from the Eastern dialect group (36%), 13 participants came from the Trøndersk dialect region (16%) and the remaining 12 participants spoke a Northern Norwegian dialect (14%). 32 reported use of both feminine and common gender (39%), 41 participants reported exclusive use of feminine gender (49%) and 10 reported exclusive use of common gender (12%).

Materials

The pretest employed a simple design that varied object length from one word (short condition) to four words (long condition) and presented all sentences in a shifted order. It was arranged as a Latin Square design: all participants saw sentences with objects of three different lengths (short, medium and long), but no item in more than one condition. I constructed 24 triplets of items. All items began with a Norwegian proper name followed by the verb, the object and the particle in the final position. The proper names were chosen from the ten most popular Norwegian baby names from the years 1977 to 1987 provided by the Norwegian statistical bureau (www.ssb.no/emner/00/navn/). This range was chosen to reflect names that are common among people of similar age as the intended participants or slightly older. Half of the names were feminine names. If names had too much orthographical overlap (e.g. Katrin and Katrine) only one name was chosen to avoid memory effects, if one actor seems to appear twice in the same task performing different actions. If names were similar to a German name of the opposite sex (e.g. Sigfrid, a female name in Norwegian, but a male one in German) or were perceived as too unfamiliar to L2 learners, they were excluded in order to avoid confusions regarding the actor of the sentence that would lead to increased regarding times for the proper name. The particle verbs were chosen from the two most common German text books of Norwegian and randomly assigned to the proper names. The verbs and especially the particles that went along with them were chosen to represent the frequency of occurrence in spoken language. Five particles were used four times (*ned* ‘down’, *bort* ‘away’, *inn* ‘in’, *ut* ‘out’ and *opp* ‘up’), two other particles were used twice (*på* ‘on’, *av* ’off’). In most cases the particle adds an aspectual or nuanced meaning to the original meaning of the verb (e.g. *spise – spise opp* ‘eat – eat up, *ta – ta av* ‘take – take off’). In four cases the particle changes the original meaning of the verb to a new, more idiomatic meaning: *pusse – pusse opp* ‘clean – renovate, *si – si opp* ‘say – terminate’, *slå - slå på* ’hit – turn on’, *slippe – slippe inn* ’slip, let go - let in’. With the exception of *pusse opp* and *si opp* all the German translations of the Norwegian particle verbs are particle verbs as well. The length of the shifted object increased from one word (a definite noun) in the short condition to three words (a definite noun with one adjective) in the mid-length condition to four words (a definite noun with two adjectives) in the long condition. The length of the adjectives and nouns varied between the individual items and if length is treated as a continuous variable it varies between two and fourteen syllables. An overview over the different lengths of the objects in syllables and letters can be found in Appendix A.

Short objects had an average length of 7.25 letters (range: 5-13) or 2.5 syllables (range: 2-5), mid-length objects had an average length of 18.38 letters (range: 12-26) or 6.88 syllables and long object had an average length of 23.63

letters (range: 18-37) or 9.12 syllables. Length in letters and number of syllables are correlated, but they are less indicative of which condition the item comes from as the same number of syllables or letters can be found in different conditions, i.e. an item with ten syllables can be part of the mid-length or the long condition, as can be an item with twenty letters.

The long condition with four words might not seem long compared to the objects of eight words or more found in corpus studies (e.g. Lohse et al., 2004) and it would have been possible to further increase the length of the object. However, I wanted to keep the structure of the objects as simple as possible and create the most natural-sounding sentences possible. Therefore I did not introduce relative clauses that modify the NP to increase the length of the object as this would also increase syntactic complexity. The possibility of adding extra adjectives to increase object length was also rejected as unnatural-sounding. Using coordinated NPs as the object – e.g. 'lets the dog and the cat in' – could also have caused processing difficulties because the 'and' could also introduce a new main clause, making these objects less comparable to shorter objects with only one entity and no additional ambiguity. As both the EIC and distance-based processing accounts predict a decline of acceptability and processing ease already for shorter and less complex objects, I decided that using objects consisting of four words keeps the sentences natural, while ideally showing effects of object length and was therefore the best option. A sample set of items used in the pretest is given below. The object is printed in bold. A full list of all items can be found in Appendix A.

short object
(119a) Anders slipper **hunden** inn.
Anders lets dog-the in.
'Anders lets the dog in.'

mid-length object
(119b) Anders slipper **den blinde hunden** inn.
'Anders lets the blind dog in.'

long object
(119c) Anders slipper **den våte, blinde hunden** inn.
'Anders lets the wet, blind dog in.'

The 24 items were spread across three lists using a Latin square design. Twelve fillers were added that contained a case of ungrammatical object shift of a short object across the negation. The order of items and fillers was pseudorandomized across the entire list.

Procedure

The platform, distribution and presentation used in this pretest were the same as described under Section 4.2. The task in this pretest was an acceptability rating. Participants were asked to assign each sentence a rating on a Likert scale from 1 to 5 with 1 denoting highly acceptable sentences and 5 unacceptable sentences. I asked for acceptability and not grammaticality as all experimental sentences are in theory grammatical, but are expected to vary in acceptability. Acceptable was therefore defined as grammatically correct and naturally sounding. The completion of the questionnaire took on average less than 10 minutes.

Results

The data of three participants was excluded after visual inspection of the data as they had rated the ungrammatical fillers as acceptable (i.e. average ratings of 1.0, 2.17 and 2.75 compared to an average of 4.6 for the rest of the group) and they therefore either did not recognize the ungrammaticality or had misinterpreted the rating scale. After removing these participants, the fillers received the intended rating as ungrammatical/ inacceptable with an average of 4.6 (SD: 0.5). The results presented in this discussion are from the analysis of the remaining 80 datasets.

As can be seen from Table 7.3 below, the rating for the shifted objects gets higher the longer the object is, but even the condition with the long object received a rating that is still within the acceptable range, indicating that it was not perceived as ungrammatical.

Object type	**M**	**SD**	**range**
short	2.59	0.74	1.42 – 4.23
medium	2.80	0.83	1.58 – 4.26
long	2.93	0.78	1.89 – 4.15

Table 7.3 Mean item ratings across conditions with standard deviations

As can be seen in Table 7.3, the highest ratings (4.23, 4.26, and 4.15) in each length condition differed by more than one point from the overall average across all three conditions. It was the same item that contributed this high average in all three conditions. Its removal from the dataset affected the average ratings only minimally (2.51, 2.73 and 2.87 respectively). The overall rating pattern was unaffected by the removal and the item was flagged in all subsequent analyses and the self-paced reading experiment reported in Section 7.4. The ANOVA used a categorical definition of length and revealed a main effect of Condition ($F_1(2, 158)=19.14$, $p<0.001$, $F_2(2,46)=7.23$, $p=0.002$). Post-hoc comparisons showed that all three ratings were significantly different from each other in the by-subjects comparisons. Short objects were significantly more acceptable than medium-length objects ($t_1(79)=-3.84$, $p<0.001$, $t_2(23)=-2.28$, $p=0.03$) and long objects ($t_1(79)=-5.48$, $p<0.001$, $t_2(23)=-3.35$, $p=0.003$). The difference between medium-length and long objects only turned out significant in the by-subjects comparison ($t_1(79)=-2.71$, $p=0.008$, $t_2(23)=-1.758$, $p=0.09$). ANOVAs run on the rating with gender system, dialect or written language variety as between-subjects factors did not show any significant effects of any of these factors on the ratings ($ps>0.5$).

The weaker results in the by-item comparisons become clearer when looking at the ratings for selected items across conditions as presented in Table 7.4. A full list of all ratings can be found in Appendix A. The overall trend of longer objects receiving worse ratings is not borne out for all items equally and the trend has different strengths and endpoints. Some items, like 16, remain in the acceptable range in all three conditions, while others (e.g. item 21) reach the unacceptable range only in the long condition. Other items even show a reverse trend and were rated as more acceptable in the long condition (e.g. item 19) or show a spike in acceptability in the mid-length condition (item 13). Some items also show little to no change in acceptability across conditions (items 22 and 23). Item 22 received generally unacceptable ratings and will be flagged in the self-paced reading experiment as it might be perceived as ungrammatical during reading or cause disproportionate processing difficulties compared to other items.

Item	Short	Mid	Long
13	2.31 (1.44)	3.33 (1.33)	2.96 (1.34)
16	1.65 (0.85)	2 (1.07)	2.3 (1.03)
19	2.31 (1.41)	2.52 (1.05)	1.96 (0.85)
21	2.89 (1.34)	3.85 (1.16)	4 (1)
22	4.23 (1.21)	4.26 (0.81)	4.15 (0.99)
23	3.78 (1.12)	3.74 (1.13)	3.73 (1.08)

Table 7.4 Selection of item ratings, SD in brackets

Figures 7.1 and 7.2 plot the mean ratings based on the number of syllables and letters of all experimental items. They present object length as a continuous variable rather than a categorical one with three levels. Both plots show the general trend of an increase in average rating and therefore decreasing acceptability. Linear regression performed for either factor showed that both are significant (both $ps<0.001$) predictors of average ratings and explain comparable amount of data (adjusted R^2 for syllables = 0.034, adjusted R^2 for letters = 0.032). The plots do however show that there is more variation in the letter plot than in the syllable plot. Both plots show the same high outlier at the end of the scale which is the flagged item that received a very unacceptable rating. There is also another low outlier with 12 letters or four syllables that received unusually good ratings. This outlier consists of data from two items – one from the short condition and one from the mid-length condition.

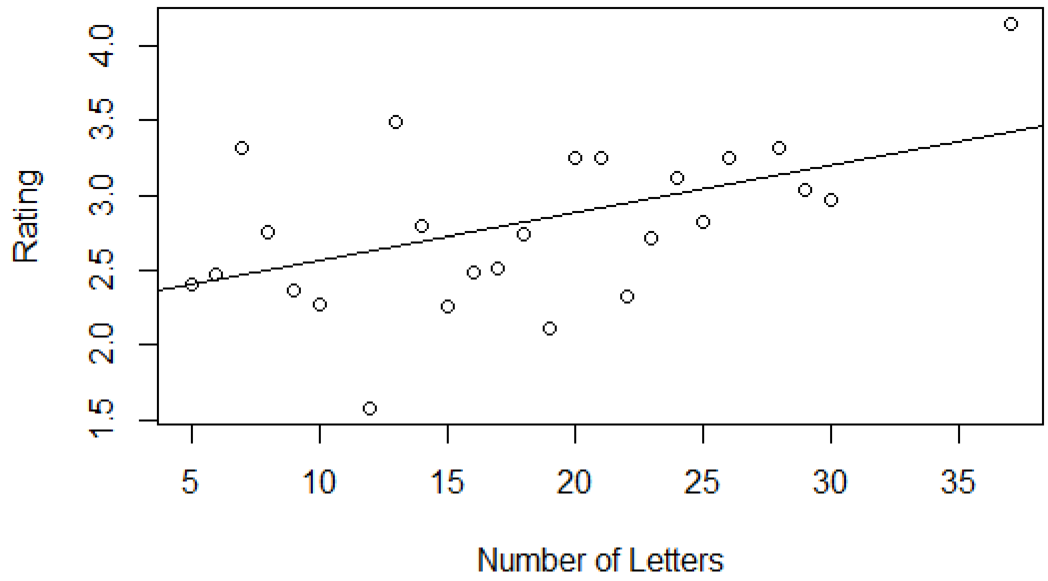

Figure 7.1 Mean ratings plotted against the number of letters

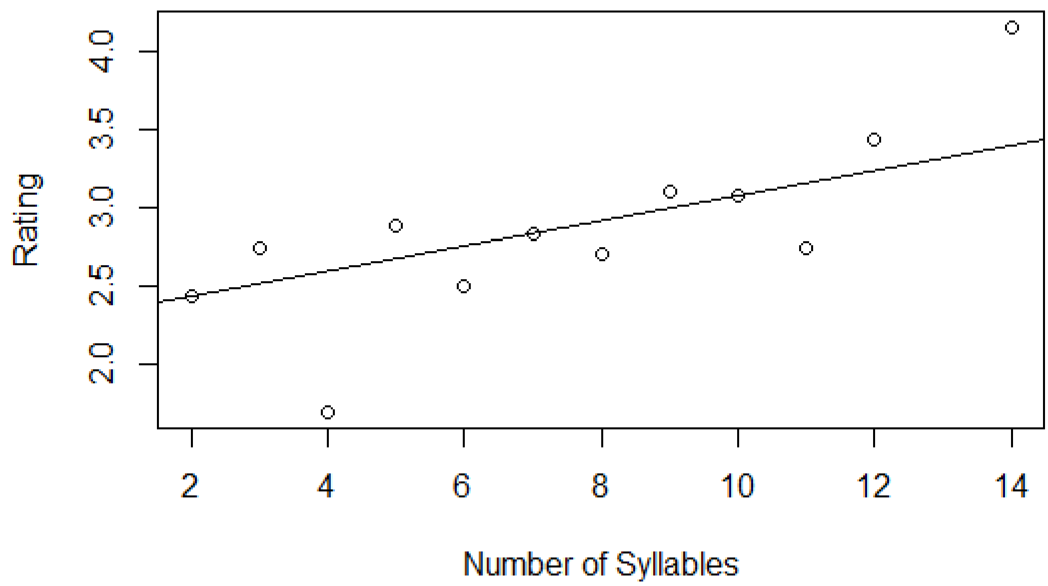

Figure 7.2 Mean ratings plotted against the number of syllables

Discussion

The main aim of this pretest was to identify whether objects consisting of four words were still acceptable in a shifted position and to decide between choosing mid-length or long objects for the self-paced reading task. As objects in the longest condition were still deemed acceptable, the short and the long condition were chosen for inclusion in the self-paced reading task.

The ratings for shifted object constructions also got significantly worse with increasing number of letters and syllables, and both seemed to be equally good predictors for the rating outcome. This is in line with descriptive grammar that assumes a decline in acceptability for shifted objects with increasing object weight. As number of letters and syllables were significant predictors in separate linear regressions, they will also be included in the post-hoc analysis of the acceptability rating task of Experiments 3a and 3b.

7.3 Experiment 3a: Acceptability rating task

This task was administered to check whether the L1 and L2 participants had any preference for any of the two object orders (particle-first, object-first) and whether this preference was influenced by the length of the object. It was also supposed to examine the participants' knowledge of the grammaticality of various types of object shift. This task also served the purpose of verifying whether L2 speakers have both orders of particle and object available in their interlanguage grammar.

Participants

The participants in this experiment were the same as described in section 4.3 with 32 participants in each group. L1 German speakers were chosen as the L2 group as their native language also has particles that can be split from verb stems, but the position of this particle is always at the end of sentence. So while the L2 group is familiar with the structure itself, their native language provides them with only one grammatical option as opposed to two that are, in theory, equally valid options in the L2 investigated in this experiment.

Materials

This experiment had a simple design with the factor Order and its two levels (object-first vs. particle-first). The particle-first order was added to the object-first order that was used in the pretest. Length was a between-items factor: any item was either short (120a+b) or mid-length (121a+b). Mid-length objects were used in this task instead of long objects as in the self-paced reading task in order to avoid too much repetition. There were 16 critical sentences in this task. A subset of the given names used in the pretest was used in this task. The verbs from the pilot study were replaced with other particle verbs half of them expressed the opposite action or a related task as the verbs used in the pilot study. This was also done to avoid repetition as the verbs from the pilot study were used in the self-paced reading experiment. As there were only short and mid-length objects, the range in number of letters and number of syllables was smaller than in the pretest, 5-26 and 2-9 respectively.

Short, object-first (OP)
(120a) Stian slår lyset av.
'Stian turns the light off.'

Short, particle-first (PO)
(120b) Stian slår av lyset.
'Stian turns off the light.'

Mid-length, object-first (OP)
(121a) Ingrid slipper den temperamentsfulle hesten ut.
'Ingrid lets the spirited stallion out.'

Mid-length, particle-first (PO)
(121b) Ingrid slipper ut den temperamentsfulle hesten.
'Ingrid lets out the spirited stallion.'

There were also 16 fillers, ten of which were related to the experimental sentences in that they contained subject or object shift. Four sentences with subject shift were all grammatical. Of the six sentences with object shift, only two were grammatical. The remaining six sentences were all ungrammatical sentences with ungrammatical reflexive verbs, ungrammatical particle verbs or incorrect past tenses. Altogether, the participants had to rate 22 grammatical sentences and 10 ungrammatical sentences. This is to say that I purposefully deviated from the usual 50:50 ratio of grammatical and ungrammatical sentences. Particle shift and subject shift are optional variations of each other given a correct representation in the speaker's grammar and could not be constructed using an ungrammatical shift. All sixteen experimental items were therefore grammatical in either condition. In order to avoid a bias and make the task too obvious for the participants by having only ungrammatical fillers, I reduced the number of ungrammatical fillers. Increasing the overall number of items in this task to reach a 50:50 ratio, while still hiding the experimental items among grammatical fillers would have necessitated the addition of perhaps 12 ungrammatical fillers putting the overall number of items in this task at 44. As this experiment was part of a larger experimental session (see procedure in Section 4.3), adding this many additional items would have likely exhausted participants. It also not clear whether all L2 speakers have a correct representation of the various shifts in their L2 grammar, and assessment of the knowledge of the grammatical rules was an aim of this task. The ratio of grammatical and ungrammatical sentences for individual participants, especially in the L2 group, might have been very different from the 22:10 ratio that was intended as acceptability of shift has been found to vary by participants.

Procedure

The acceptability rating task was an untimed pen-and-paper questionnaire. The instructions told the participants that they should judge the acceptability of the sentences to follow. Acceptability was defined as sounding natural in an everyday reading. Participants were asked to give their rating on a Likert Scale from 1 until 5, with 1 being acceptable and 5 being unacceptable. Two examples were given to illustrate the end points of the scale. The acceptable example was the sentence *Anne leser ikke boken* ('Anne does not read the book') in which the object has correctly remained in situ. The unacceptable sentence **Boken være spennende*' (*The book be exciting') contained a verb that had not been conjugated for tense. The unacceptable sentence did not contain an incorrect shift in order to avoid focusing participants' attention on shifted objects and therefore bias them towards one order. An ungrammatical version of the acceptable sentence (for instance, **Anne leser boken ikke*) could have biased especially the L2 group towards unshifted objects. Participants were also told that they were free to indicate the source of the unacceptability in the sentence and there was enough space in the questionnaire to mark the error or write down corrections. About half of all participants made corrections to any of the sentences. The sentences were written in Verdana font size 8. The ranking scale was shaded and the numbers were given for the participants to either circle or tick off. At the head of the scale there was a reminder of which number represented which rating. The same format was used as in the acceptability rating task reported in Section 5.2 (see Figure 2.4 for an example). The critical sentences were spread across two lists using a Latin Square design. Two additional lists presented the items in a reversed order.

Predictions

The Norwegian Reference Grammar states that both orders of particle and NP object are equally acceptable and even though there are some Norwegian dialects (in Trøndelag and Eastern Østlandet) that only feature the particle-first order, a general preference based on dialect is not to be expected in this experiment as none of the native speakers were recruited from any of these dialectal regions. As both orders are grammatical, the average rating of the shifted order should not reflect an evaluation as ungrammatical (i.e. an average beyond 4).

The predictions are based on the manipulation of the particle and object order. An additional effect of object length in the shifted orders is predicted based on the results of the pilot study, although object length was only a between-items factor in this experiment.

A – If the rules stated by descriptive grammar are applied by the participants, then there should be full optionality for short objects with acceptable ratings in both orders. For mid-length objects the shifted order should receive less acceptable ratings than the unshifted order.

B – If there is a general preference for the more frequent order, the unshifted order should get better ratings than the shifted order as this is by far the more frequent order in corpus studies.

C – If the influence of the object length is more gradient, we should see a gradual decrease in acceptability for shifted objects based on the number of syllables or letters of the object that was also found in the pilot study.

L2 predictions

D – As the shifted order is a valid option in Norwegian and the only grammatical order in German, speakers might simply prefer this order and continue to use it also in their L2 (see the German L2 group in Bohnacker's 2007 study). If this is the case, the shifted order should receive a lower average rating than the unshifted order, independent of the length of the object.

E – Similar to the frequency-acceptability paradox in German scrambling, L2 speakers might not have learned that both orders are grammatical and acceptable, especially as the shifted order is very infrequent. If L2 speakers have a faulty L2 grammar motivated by the differences in frequency of the two orders, they might rate the shifted order as highly unacceptable, if not ungrammatical.

	short-PO	**short-OP**	**mid-PO**	**mid-OP**
A – descriptive grammar rules	1	1	1	2.5
B – frequency	1	2	1	2
C – object length	1	1	1	1.5 – 3
D – German order preference	2	1	2	1
E – faulty L2 grammar	1	5	1	5

Table 7.5 Idealized judgment patterns based on the predictions

Results

Main analysis

The general sensitivity of participants to violations of shift patterns was assessed through their performance on the fillers featuring object and subject shift. There were no participants who were unable to identify ungrammatical structures, or who consistently judged grammatical structures as ungrammatical. There was also no indication that any participant had accidentally swapped the scale or not paid attention to the task. Therefore, no participant was removed from the dataset. The full dataset included 1024 observations.

Table 7.6 below shows the overall rating averages and separate averages for the L1 and L2 groups by experimental condition. The particle-first order is generally rated more favorably than the object-first order. Length in the object-first condition seems to matter more to the L1 speakers who rate the short items in the object-first condition less favorably than the mid-length ones, while the L2 does not seem to make a difference.

	short-PO	short-OP	mid-PO	mid-OP
Overall	1.39 (0.51)	2.8 (1.07)	1.37 (0.55)	2.43 (1.08)
L1 (N=32)	1.34 (0.41)	2.74 (1.05)	1.16 (0.3)	2.03 (0.73)
L2 (N=32)	1.44 (0.6)	2.86 (1.11)	1.58 (0.66)	2.83 (1.24)

Table 7.6 Mean ratings in the acceptability judgment task, SDs in brackets

As length had not been varied for each item, it was entered as a within-factor in the by-subjects ANOVA, but as a between-factor in the by-items ANOVA. The between-groups ANOVA for this dataset showed three main effects for every factor entered into the analysis: group, order and length. There were an additional two interactions: Group x Length and an Order x Length interaction. For ease of exposition, the results of the ANOVA are displayed in Table 7.7 below.

Effect		**F-value**	**p-value**	**significance**
Group	(1,62) (1,14)	11.26 7.74	0.0014 0.015	** *
Order	(1,62) (1,14)	64.44 105.15	<0.001 <0.001	*** ***
Length	(1,62) (1,14)	10.89 0.83	0.0016 0.38	** n.s
Group x Order	(1,62) (1,14)	0.42 1.33	0.52 0.27	n.s n.s
Group x Length	(1,62) (1,14)	17.92 3.73	<0.001 0.074	*** n.s
Order x Length	(1,62) (1,14)	8.01 2.17	0.006 0.16	** n.s.
Group x Order x Length	(1,62) (1,14)	2.14 1.13	0.15 0.31	n.s. n.s.

Table 7.7 Results of the between-groups ANOVA for the full rating dataset

Based on the Group x Length interaction, I analyzed the L1 and the L2 data separately to explore the effect of Length in the individual groups. The L1 group showed main effects of Order ($F_1(1,31)=48.00$, $p<0.001$, $F_2(1,14)=46.34$, $p<0.001$) and Length ($F_1(1,31)=30.62$, $p<0.001$, ($F_2(1,14)=2.27$, $p=0.15$) and an Order x Length interaction ($F_1(1,31)=9.49$, $p=0.004$, ($F_2(1,14)=2.6$, $p=0.13$). An exploration of the interaction through t-tests showed that this interaction was caused by a smaller rating difference in the mid-length items ($t_1(31)=6.31$, $p<0.001$, $t_2(7)=4.66$, $p=0.002$) than in the short condition ($t_1(31)=6.27$, $p<0.001$, $t_2(7)=5.07$, $p=0.001$). The L2 group only showed a main effect of Order ($F_1(1,31)=26.26$, $p<0.001$, $F_2(1,14)=110.82$, $p<0.001$), all other Fs<1. This suggests that the Group x Length interaction found in the between-groups ANOVA reflects an effect of Length that is restricted to the L1 group.

Post-hoc analysis

Inspection of the ratings for individual items showed one item (number 23) that had been judged as fairly unacceptable by the native speakers in both conditions (PO condition M=3.3, OP condition M=4.06). The rating in the PO condition marked this item as an extreme outlier compared to the other items in this condition and suggested an issue with the item independent of the experimental manipulation. In order to see whether this had an effect on the results, I re-ran the analysis with this one item removed from the dataset of both groups. This caused the removal of 64 data points (6.25%). As item 23 was a short item, there are only seven items with short objects left in the analysis, opposed to eight items with mid-length objects.

The removal of this one item resulted in a change of the ratings for the short condition, whereas the mid-length ratings remained unaffected. The ratings of the L1 group are affected more strongly as they had perceived the item as nearly ungrammatical, while the L2 group had given it a rating similar to the other items. The rating for the particle-first order now shows a floor effect in the L1 group with an almost perfect 1.05 rating as seen in Table 7.8.

	short-PO	short-OP
Overall	1.28 (0.53)	2.75 (1.11)
L1 (N=32)	1.05 (0.14)	2.63 (0.98)
L2 (N=32)	1.5 (0.66)	2.88 (1.23)

Table 7.8 Average acceptability ratings in the short condition after removal of one item

The removal of the item also affected the results of the ANOVA as can be seen when comparing Table 7.9 to the previous analysis in Table 7.7.

Effect		**F-value**	**p-value**	**significance**
Group	(1,62) (1,13)	18.68 52.62	<0.001 <0.001	*** ***
Order	(1,62) (1,13)	65.83 89.87	<0.001 <0.001	*** ***
Length	(1,62) (1,13)	3.6 0.14	0.062 0.7	n.s. n.s.
Group x Order	(1,62) (1,13)	0.007 0.64	0.79 0.44	n.s. n.s
Group x Length	(1,62) (1,13)	4.47 4.12	0.038 0.062	* n.s.
Order x Length	(1,62) (1,13)	15.1 1.9	<0.001 0.19	*** n.s.
Group x Order x Length	(1,62) (1,13)	7.69 3.11	0.007 0.1	** n.s.

Table 7.9 Results of the between-groups ANOVA after the removal of one item

The main effect of Length is now only marginally significant, the effect of the Group x Length interaction decreases and a Group x Order x Length interaction appears. The two interactions with Group justified a split of the L1 and L2 data. The L1 group showed the same effects as in the full analysis. main effects of Order ($F_1(1,31)=75.53$, $p<0.001$, $F_2(1,13)=45.8$, $p<0.001$) and Length ($F_1(1,31)=8.32$, $p=0.007$, ($F_2(1,13)=1.21$, $p=0.29$) and an Order x Length interaction ($F_1(1,31)=30.00$, $p<0.001$, ($F_2(1,13)=3.37$, $p=0.089$). The main effect of Length was smaller than in the full analysis, while the effect of the interaction had gotten bigger. The exploration of the interaction showed the same effect as in the full analysis, except that the effect of Order had grown even bigger in the short condition due to the floor rating in the particle-first condition ($t_1(31)=9.23$, $p<0.001$, $t_2(6)=4.98$, $p=0.0025$). The results of the t-test in the mid-length condition did not change as no data had been removed from this condition. The rating difference between the two order conditions is bigger in the short condition (1.6) than in the mid-length condition (0.87). Just as with the full dataset, the L2 group only showed a main effect of Order ($F_1(1,31)=22.07$, $p<0.001$, $F_2(1,13)=103.00$, $p<0.001$), all other Fs<1.

In order to investigate a possible gradient effect of object length on the acceptability of the OP order in the L1 group, I ran separate linear regressions with length in letters and length in syllables as continuous predictors. These analyses yielded no significant result (ps>0.3).

The ratings of the L2 group showed a lot more variation than in the L1 group and merit closer inspection. The highest ratings for the OP order in the L1 group were between 3.5 and 4, and the biggest difference between the PO and OP averages in any L1 participant was 2.5. Five L1 participants also showed a better acceptability rating for the OP order compared to the PO order (a difference of up to 0.625). Two participants in the L2 group rated the OP order as completely ungrammatical (average of 5) and another three had averages of 4.25, 4.5 and 4.87 exceeding the maximum rating of any L1 participant. Some L2 ratings in the PO order also exceeded the maximum L1 ratings (average of 2) up to an average of 3.25. The bigger number of more extreme ratings in both conditions also leads to more pronounced rating differences in the L2 group. The biggest advantage for the OP order reached up to 2.125 points and the advantage for the PO order extended to 4. As the L2 group was very homogeneous with regard to L2-specific variables such as AoA, proficiency and exposure to Norwegian, no additional analyses with these variables were conducted.

Discussion

L1 and L2 groups showed different results in the acceptability rating task. The L1 group showed effects of order and object length that were not predicted by any of the idealized ratings above as the effect of length went into the opposite direction as assumed by prediction C, with more acceptable ratings for longer objects compared to shorter objects in the shifted condition. The result of the L2 group mirrors prediction B with a general preference for the more frequent unshifted order in which verb and particle are adjacent.

The results of the L1 group in this task deviate from the claims made by the Norwegian Reference Grammar in two points. One is the claim that at least with short objects, both orders are optional variations of each other. There is no support for this claim in the acceptability data as there was a significant difference between the two ratings in the short condition and the shifted order was perceived as a lot less acceptable than the unshifted order. The second claim regards the decrease of acceptability for the shifted order with increasing object length. Length was only a between-items factor in this experiment which might weaken the result, but the effect was nevertheless significant and it went in the opposite direction from the expected effect based on the Norwegian Reference

Grammar. The acceptability of the shifted order increased from the short to the mid-length condition. This result not only goes against the claim of prescriptive grammar, it is also at odds with the results from the pilot study in which a general trend with decreasing acceptability had been found across the three conditions employed in the pretest and in more specific post-hoc analyses using number of letters or syllables as predictors. No such effect of number of letters or syllables had been found in this experiment. The failure to reproduce the results from the pretest could be explained by the fact that different items and subjects were used in the two tests: preferences for either order could be subject-specific as well as item-specific as had been already suggested for regular object shift by Josefsson (2010). Experiment 3a also used fewer items and fewer participants than the pilot study leading to less statistical power. It is possible that the acceptability of a shifted object is evaluated on an item-specific basis that considers other factors in addition to the pure number of letters or syllables of the object, such as for example the actual verb used or the type of the object, as has been suggested for example by Gries (2002).

The L2 group only showed sensitivity to the shift manipulation, but no sensitivity at all to changes of object length. The L2 group was also generally stricter in their ratings which resulted in overall higher averages. This result suggests that the L2 speakers were aware of the fact that both orders are acceptable, but clearly preferred the order that is more frequent in the target language. However, some participants showed extreme preferences for one order over the other. One subgroup showed the pattern of prediction E judging the shifted order as completely unacceptable (averages of 5). A complete rejection of the shifted order was not found at all in the native group. This L2-specific behavior could be caused by the low frequency of the shifted order or by a possible perception of the shifted order as being 'too German'. For these participants, the L2 interlanguage grammar only allows the unshifted order. L2-specific factors had no influence on the ratings and extreme ratings such as the rejection of the shifted order cannot be explained by lack of proficiency. The L2 participants should know that either order is acceptable based on their proficiency and the time they have been exposed to Norwegian either in class or in the country. They had all learned Norwegian for at least a year and most had spent a semester at a Norwegian university. On a somewhat worrisome note, one of the participants that rated the shifted order as completely ungrammatical was also a teacher of Norwegian as a foreign language who probably also taught this structure as ungrammatical to her students thereby introducing a non-native behavioral pattern. Speakers who fail to identify the shifted order as grammatical will most likely not get corrective feedback from native speakers as the production of the unshifted order is in line with the preference found in the L1 group.

There was also a subgroup that showed the opposite pattern with a preference for the shifted order, i.e. the order that is grammatical in German, corresponding to prediction D. This preference was less extreme than the rejection in the other subgroup just reported. It could reflect an influence of the native grammar that makes up for the lower frequency of the shifted order in the L2. When looking at the frequency of the two orders independent of the language in question, the L2 group probably has more experience with the shifted order due to their native language. Bohnacker (2007) had also found prolonged use of the German order in L2 Swedish in which the shifted order is ungrammatical. In the absence of a second L2 group with a different order preference or no particles at all in their L1, the role of transfer remains inconclusive. Anecdotal evidence from one of the native speakers also suggested that the shifted order can be a possible hallmark of a German accent in Norwegian. The native speaker had remarked that the shifted constructions sounded a lot like her husband [a German] when he spoke Norwegian. Unlike the complete rejection of the shifted order as in the previous subgroup, a preference for the shifted order therefore seems to be more noticeable for native speakers and will be more likely to receive corrective feedback.

The overall preference for the unshifted order that was found in both groups is in line with a syntactic analysis that takes the particle > object order as the canonical order. However, it also reflects the higher frequency of this order that does not necessarily depend on a lower syntactic complexity. Finally, the preference for the particle > object order is also in line with the EIC that makes no claims regarding syntactic differences between the two orders.

When comparing the results of the two acceptability rating tasks of experiment 2a and 3a, we see that the L1 groups always made a significant difference between the two orders presented. This was not the case for the L2 group. The L2 group in experiment 2a on German scrambling rated both orders as equally acceptable and failed to perceive a difference in applicability in context-free situations. The L2 group in experiment 3a was sensitive to a difference in acceptability between the two orders. The difference between the two orders is rather subtle in both experiments and does not cause ungrammatical sentences or changes to the sentence's content. The task was also exactly the same down to wording of the instructions, assuming that L2 speakers with a Slavic background and a German background comply with instructions and understand acceptability in the same way, there should also be no task effect. What then accounts for the differences? Why is there a native-like pattern in one experiment, but not the other? I suggest it is due to the way the two orders are differentiated from each other. The speaker in experiment 2a has to correctly process and assign case to recognize the difference between the sentences, but it is purely surface word order that signals the difference in experiment 3a. When considering the EIC and the constituent-to-word ratio in the two experiments, there is no difference between the conditions with regard to the ratio in experiment 2a, but there is one condition that has a more favorable constituent-to-word ratio in experiment 3a. This turns out to be the more acceptable order also for the L2 speakers. It is possible that the application of a mechanism reflecting language-independent processing advantages such as the EIC helps L2 speakers in identifying ordering differences.

7.4 Experiment 3b: Self-paced reading task

Participants

The participants were the same as described in section 4.3 with 32 participants in each group.

Materials

The 24 experimental sentences used in this self-paced reading task were based on the materials used in the pilot study (Section 7.2). Names, verbs and objects were reused. The mid-length condition was dropped to test the biggest difference in object length. This drop also increased the number of data points per condition, as there were now six data points per condition collected per participant. A second order condition was introduced in which the particle followed right after the verb. This is the unshifted position of the object. All experimental sentences were extended by adding an adverbial consisting of two words. This addition moved the critical region away from the end of the sentence and made it possible to check spillover effects on the first word after the critical region.

The self-paced reading task employed a 2x2 design with object shift (object-first vs. particle-first) and object length (short vs. long) as factors. In the following examples, the object is printed in bold.

Short, object-first (OP)
(122a) Anders slipper **hunden** inn av medlidenhet.
Anders lets dog-the in out of compassion
'Anders lets the dog in out of compassion.'

Short, particle-first (PO)
(122b) Anders slipper inn **hunden** av medlidenhet.
'Anders lets in the dog out of compassion.'

Long, object-first (OP)
(122c) Anders slipper **den våte, blinde hunden** inn av medlidenhet.
Anders lets the wet, blind dog-the in out of compassion.
'Anders lets the wet, blind dog in out of compassion.'

Long, particle-first (PO)
(122d) Anders slipper inn **den våte, blinde hunden** av medlidenhet.
'Anders lets in the wet, blind dog out of compassion.'

As with the materials in study 1, words with a frequency below 100 occurrences in the Oslo corpus of tagged texts were included in the vocabulary list to check whether they were understood correctly by the participants.

Participants completed 72 experimental trials, 24 with items from this experiment, 24 with items with items from Experiment 1b and 24 fillers. Participants also answered 24 comprehension questions. Eight of the questions targeted items from this experiment and probed general sentence comprehension without direct or indirect reference to the experimental manipulations. Of the eight comprehension questions on experimental items from this experiment, half expected a negative answer. The remaining 16 comprehensions questions followed items from Experiment 1b or fillers. There was a short practice session with four trials that were unrelated to the experiment and one comprehension question.

Randomization was conducted as part of the procedure described under 4.4. As the items of this experiment posed as fillers for Experiment 1b, the organization of presentation lists and fillers is the same as described for Experiment 1b.

Procedure

The procedure for the SPR task and the overall procedure of the experimental session was the same as described in Section 4.4.

Predictions

A – Based on processing theories involving ordering principles (such as the EIC) and memory-based processing theories, the PO order should be processed faster than the object-first order, because...

- ... it allows the earlier completion of the sentence's phrase structure (EIC),
- ... local-dependencies have been found to be processed faster than non-local dependencies and the placement of the object between the matrix verb and the particle creates a non-local dependency between verb and particle,
- ... it is the more frequent word order.

Additionally, for the OP order: If participants store the verb in memory in order to reactivate it when the particle is encountered, the retrieval should be more difficult after a long object than after a short one due to the presence of more intervening material. This should only affect the particle in the object-first order. The reading times of the particle in the PO condition should not show a significant difference, whereas the reading times of the particle in the OP condition should be higher in the long object condition than in the short one. Based on the performance in the acceptability rating task reported in Section 7.3, this is the expected pattern of the L1 group.

B – If the length manipulation is not big enough to create processing difficulty during the retrieval of the verb at the particle, then there should only be an effect of order with faster reading times for the PO order, and the reading times of both orders should be equally affected by the longer object. Based on the offline performance, this is the pattern expected for the L2 group as they had shown a general preference for the PO order and no influence of object length.

C – If there is no mental reactivation of the verb at the particle, then there should be no difference between the two orders and the particle should be read equally fast in all four conditions as its reading time is not affected by the non-local dependency it creates with the verb. Reading times would only be globally elevated for the long conditions.

Figure 7.3 shows the idealized reading time patterns for the entire region of manipulation, i.e. the entire object NP and the particle, and for the particle separately.

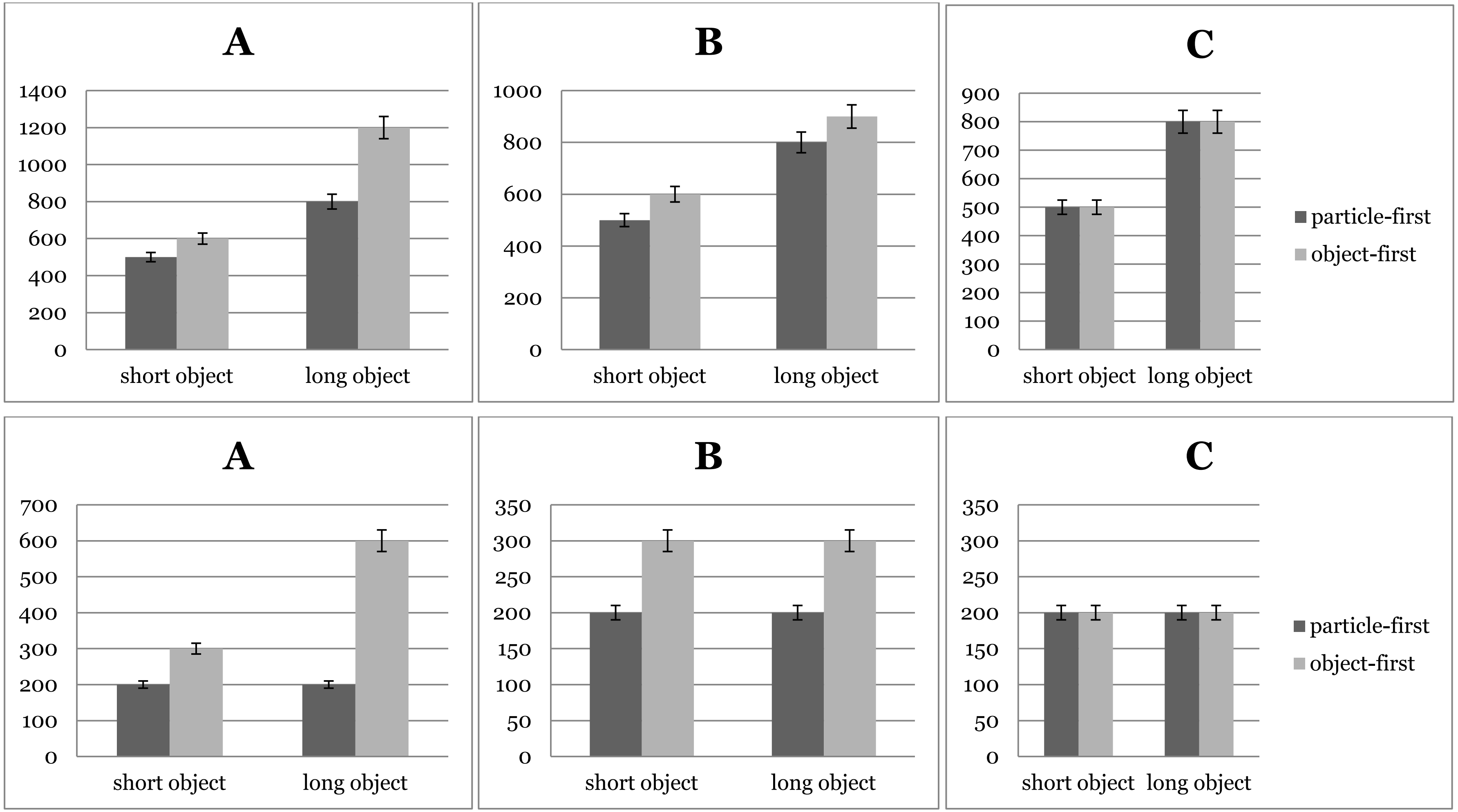

Figure 7.3 Idealized reading time patterns for Experiment 3b, upper row represents overall reading times, lower row represents particle reading time

Results

Comprehension questions

Participants had answered eight comprehension questions querying items from Experiment 3b, corresponding to a question after 1/3 of the experimental items of this experiment. A presentation of the accuracy scores for the full 24 questions asked during the self-paced reading task can be found in chapter 4.4. I repeat the accuracy scores of Experiment 3b here:

	L1 group (N=32)	L2 group (N=32)	Both groups
Experiment 3b	94.14% (0.08) (range: 75-100%)	91.8% (0.09) (range: 75-100%)	92.97% (0.09)

Table 7.10 Mean accuracy scores for comprehension questions of Experiment 3b, SDs in brackets

Unlike in Experiment 1b that had accuracy scores of around 65%, participants showed very high accuracy scores in Experiment 3b. The questions had only required a broad understanding of the experimental sentence and both participant groups were equally able to answer them. No participant was excluded from the online analysis as they had clearly all paid attention to the task.

Data cleaning – self-paced reading data

The complete dataset from the self-paced reading task consisted of 1536 experimental trials. As the number of words varied depending on the length condition, either six or nine data points were collected per sentence. All items that had been indicated as unknown in the vocabulary test by the L2 group were removed from the dataset, resulting in the removal of 23 trials (1.5%). As the short and the long condition are not directly comparable due to the length difference of the region of manipulation, they were analyzed separately. The dataset of the short condition contained 759 trials and the dataset of the long condition contained 754 trials. To prepare for the following data analysis, extreme values and outliers were removed for each segment separately. Extreme values were always identified visually by using a histogram, while outliers were identified based on a +/- 2.5 SD range around the participant mean per segment. A table with the cutoff points for extreme values and number of outliers removed for each segment can be found in Appendix B. No more than 25 data points were removed from any one segment corresponding to 3.29%.

Main analysis

Table 7.11 below shows the mean reading times per condition for all segments for both participant groups separately after the data cleaning procedure had been applied. The region of manipulation is shaded. The darker shade of grey indicates the position of the particle, in the example sentence the particle is *inn*.

As can be seen from the table above comparing the reading patterns between the groups, there seem to be no big differences between the native and non-native groups. It is clearly visible though that the L2 group is overall slower than the L1 group which should surface as a main effect of Group in the ANOVA. The region of manipulation that was analyzed contained only the NP and the particle in the short condition (two words, *hunden inn* in the example sentence in Table 7.11) and the determiner, two adjectives, the NP and the particle in the long conditions (five words, *den våte, blinde hunden inn* in Table 7.11). As the single segments varied a lot in size, i.e. the particle and the article were very short and the object noun and the adjectives were usually much longer, and were also in different positions in the two order conditions, I summed up the individual segments into one big region of manipulation for the between-groups ANOVA. Due to this difference a graphical comparison of the four conditions across the entire sentence is not very revealing as segments of different length would be compared. It can nevertheless be found in Appendix B.

The box-cox power transformation suggested a log-transformation for the majority of segments and in order to apply the same analysis to all segments, all ANOVAs were calculated on log-transformed reading times. I ran between-groups ANOVAs on the two segments preceding the manipulation that were identical across all four across conditions in order to check for possible effects preceding the actual manipulation. As expected from the reading times listed in Table 7.11 there was a main effect of Group in both segments, but no additional effects.

L1 group (N=32)	Short condition								
	Anders	slipper	hunden/inn	inn/hunden	av	medlidenhet.			
OP	507 (137)	453 (171)	474 (145)	440 (77)	404 (76)	877 (515)			
PO	526 (160)	450 (132)	411 (104)	457 (118)	395 (68)	845 (447)			
	Long condition								
	Anders	slipper	den/inn	våte,/den	blinde/våte,	hunden/blinde	inn/hunden	av	medlidenhet.
OP	542 (168)	442 (127)	414 (114)	553 (223)	528 (158)	508 (137)	430 (85)	416 (89)	915 (524)
PO	508 (149)	424 (138)	409 (115)	393 (108)	554 (221)	472 (105)	491 (130)	426 (80)	913 (563)
L2 group(N=32)	Short condition								
	Anders	slipper	hunden/inn	inn/hunden	av	medlidenhet.			
OP	633 (208)	563 (188)	688 (197)	506 (144)	465 (135)	1208 (574)			
PO	651 (245)	558 (146)	519 (173)	644 (183)	473 (99)	1249 (557)			
	Long condition								
	Anders	slipper	den/inn	våte,/den	blinde/våte,	hunden/blinde	inn/hunden	av	medlidenhet.
OP	702 (219)	551 (187)	449 (117)	777 (352)	637 (165)	672 (251)	507 (130)	455 (110)	1130 (466)
PO	729 (271)	572 (168)	491 (153)	474 (109)	811 (315)	618 (160)	677 (238)	470 (86)	1104 (579)

Table 7.11 Mean reading times for all segments across conditions for L1 and L2, SD given in brackets, darker shade indicates the particle

Before being submitted to the between-groups ANOVA, the summed data of all segments of the region of manipulation (ROM) underwent individual data cleaning according to the same criteria stated above. Sixteen data points (2.1%) were removed in the short condition and 9 data points (1.2%) in the long condition. Table 7.12 below lists the reading times for the ROM for both groups separately and collapsed over both groups. The overall average indicates the reading times that were analyzed in the between-groups ANOVA.

	OP_short	**PO_short**	**OP_long**	**PO_long**
Overall	1065 (295)	1029 (328)	2863 (892)	2787 (850)
L1 (N=32)	927 (207)	865 (218)	2519 (660)	2397 (626)
L2 (N=32)	1204 (307)	1193 (340)	3207 (969)	3177 (874)

Table 7.12 Overview of reading times of the region of manipulation in milliseconds by condition for both groups, standard deviations are given in brackets

The L1 group shows clearly faster reading times for the PO order in both length conditions (62 ms in the short condition and 122 ms in the long condition). The L2 group also shows faster reading times in the particle-first conditions as opposed to the object-first conditions, but the differences here are much smaller than in the L1 group (11 ms in the short condition and 30 ms in the long condition). The L2 group also has a lot more variation in their data as reflected by higher standard deviations that likely renders these differences not significant. The increasing object length has affected the reading times of both groups to a similar extent. The reading times of the L1 group in the long condition were 177% and 172% higher than in the short condition for the particle-first and the object-first condition respectively. The reading times of the L2 group were 166% higher in both conditions.

The following effects were found in the between-groups ANOVA on the log-transformed reading times of the ROM in the short and long condition:

Condition	Effect		F-value	p-value	Significance
short	Group	(1,62) (1,23)	20.34 155.66	<0.001 <0.001	*** ***
	Order	(1,62) (1,23)	7.66 4.89	<0.001 0.037	*** *
	Group x Order	(1,62) (1,23)	2.48 0.83	0.12 0.37	
long	Group	(1,62) (1,23)	15.88 160.22	<0.001 <0.001	*** ***
	Order	(1,62) (1,23)	2.73 1.49	0.1 0.24	
	Group x Order	(1,62) (1,23)	2.8 2.46	0.098 0.13	

Table 7.13 Results of the between-groups ANOVA on the ROM

Table 7.13 shows that there was no Group x Order interaction in either condition and only an effect of order in the short condition. The averages reported in Table 7.12 show clearly that the effect of Order in the short condition is carried by the L1 group, whereas the null result in the long condition is heavily influenced by the L2 group and the variance occurring in this group. As the two groups had shown different levels of sensitivity to the length manipulation in the offline task, and the predictions for their online processing behavior based on this offline results also foresaw different patterns, I nevertheless ran separate by-group ANOVAs, despite the absence of a Group x Order interaction in either length condition in the between-groups ANOVA. A graphical rendition of the between-groups analysis and the corresponding t-tests can be found in Appendix B. The reading time patterns found in the between-groups ANOVA reflect a mixture of the patterns found in the by-group analyses.

Separate paired t-tests for the L1 group showed a significant effect of Order in the short condition ($t_1(31)= 3.2$, $p=0.003$, $t_2(23)= 1.94$, $p=0.064$) and in the long condition ($t_1(31)= 2.9$, $p=0.007$, $t_2(23)= 1.7$, $p=0.1$). As can be seen from the averages in Table 7.12 and Figure 7.4 below, the PO order is read faster than the OP order in the corresponding length condition.

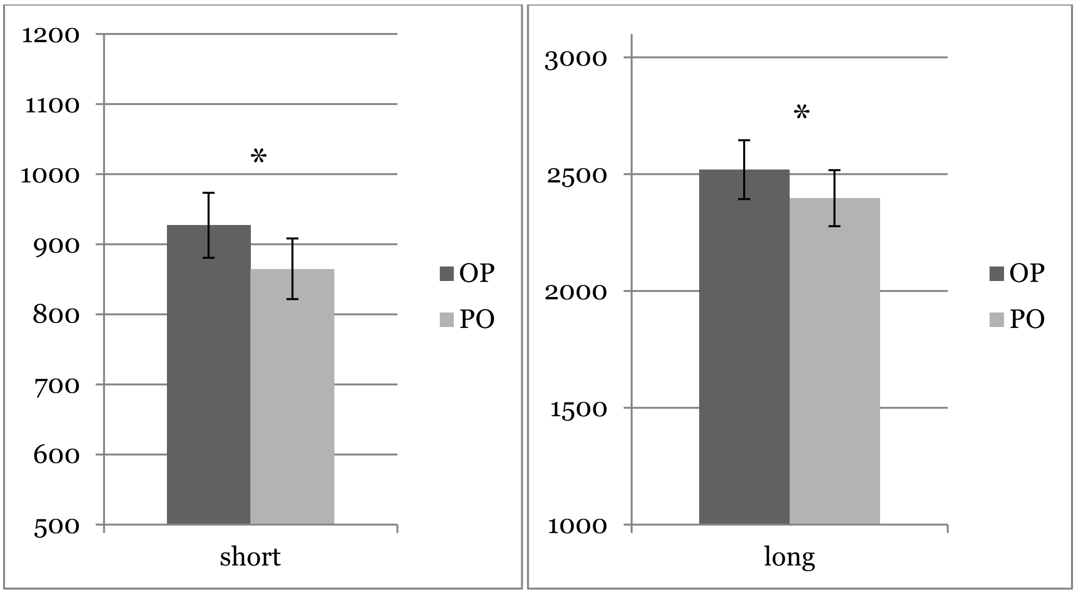

Figure 7.4 Reading times in the region of measurement per condition, L1 group

Paired t-tests for the L2 group showed no significant effect of Order in neither the short ($t_1(31)$= 0.8, p=0.42, $t_2(23)$= 1.2, p=0.24), nor the long condition ($t_1(31)$= -0.02, p=0.99, $t_2(23)$= -0.4, p=0.69). The small numerical advantage for the particle-first order did not turn out significant which is not surprising given the large variance reported in Table 7.10 and visual inspection of the graphs in Figure 7.5.

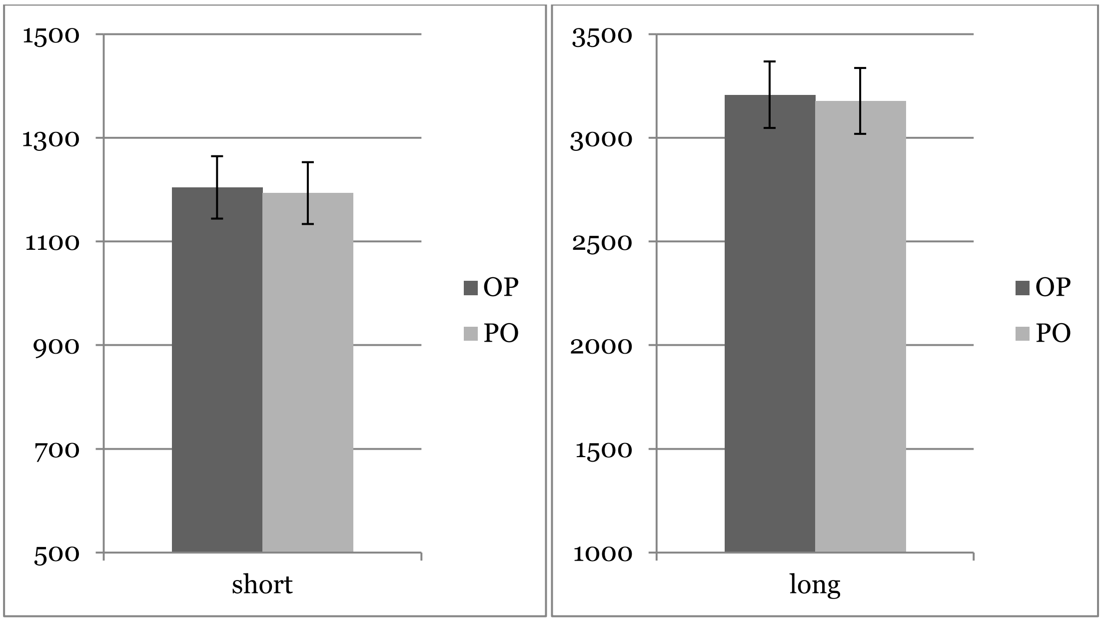

Figure 7.5 Reading times in the region of measurement per condition, L2 group

Between-groups ANOVAs run on the region following the manipulation only showed main effects of Group and a marginally significant Group x Order interaction in the by-subjects ANOVA in the long condition. There were also no effects of Order at the sentence-final region.

Post-hoc analysis

In the OP condition, the longer object could have affected the processing time of the particle more than in the short condition by creating a longer distance between the matrix verb and the particle that form a discontinuous dependency. I therefore compared log-transformed reading times of the only the particle in another between-groups ANOVA. There was again the main effect of Group ($F_1(1,62)=8.75$, $p=0.005$, $F_2(1,23)=77.4$, $p<0.001$) and additionally a main effect of Order ($F_1(1,62)=6.75$, $p=0.012$, $F_2(1,23)=7.85$, $p=0.01$) and a Group x Order interaction ($F_1(1,62)=4.00$, $p=0.049$, $F_2(1,46)=4.88$, $p=0.037$).

	OP_short	**PO_short**	**OP_long**	**PO_long**
L1 (N=32)	433 (78)	402 (101)	425 (86)	400 (100)
L2 (N=32)	495 (141)	495 (154)	486 (115)	489 (154)

Table 7.14 Average reading times for the particle per condition per group

Separate by-groups ANOVAs showed a main effect of Order in the L1 group ($F_1(1,62)=8.02$, $p=0.008$, $F_2(1,23)=10.65$, $p=0.003$) that is caused by faster reading times for the particle in the PO conditions. The reading times of the particle in the OP condition were not influenced by the length manipulation. There were no effects in the L2 group, all $Fs<1$.

One item had been flagged due to its rating results in the pilot study. All analyses were also performed with this item excluded and showed the same effects. In order not to weaken statistical power and as the participants in the SPR task differed from those in the pilot study, the item was kept in the analysis.

As only the L1 group showed any kind of effect of length, I used their data to investigate possible effects of syllable number and letter number of the object on the processing of the ROM that extend beyond the broad short/long distinction. Neither of the two linear regressions showed a significant result (ps>0.23). The trend in the data went towards an increasing difference between the two order variations with increasing length measured by letters or syllables.

Discussion

The L1 and L2 groups showed different reading time patterns in the SPR task. The L1 group showed shorter reading times for the PO order in both length conditions and no increase of reading time for the particle based on the length manipulation. This pattern corresponds to prediction B. The L2 group did not show an effect of the order manipulation nor of the length manipulation as both orders were processed equally fast in both length conditions, and the particle has nearly identical reading times across all four conditions. This separate analysis was not supported by a Group x Order interaction in the overall ANOVA. However, given the results of the separate by-groups ANOVAs, the null result in the L2 group is not enough to cause an interaction in the overall ANOVA. The high variance found in the L2 data also influenced the outcome of the overall ANOVA more strongly than the L1 data, especially in the long condition. The overall ANOVA is therefore a mixture of the two different patterns found in the two participants groups.

If we interpret the result of the overall ANOVA, we find that it does not fit with any of the predictions. The presence of faster reading times for the PO order in the short condition, but not in the long condition seems to contradict the assumptions of the EIC. In the short condition, both orders have the same constituent-to-word ratio that would not motivate a processing advantage of one order over the other. Frequency of occurrence of the PO order and the avoidance of a non-local dependency in this order could still facilitate processing of the PO order. In the long condition, the PO order has a much more favorable constituent-to-word ratio compared to the OP order that requires the processing of the full object NP before the particle verb can be identified as such. The lower frequency of the OP order and the presence of a non-local dependency that stretches across more words than in the short condition additionally disfavor the OP order. Yet, the result of the overall analysis suggests that in the long condition these three factors do not lead to a measurable slowdown in the OP order compared to the PO order. How can this be explained? The predictive value of the EIC might depend on the method used. The theory was established based on corpus data and previous studies using an SPR paradigm have not found evidence for its predictions (Konieczny, 2000). The corpus data for Norwegian only gives a general higher frequency for the PO order, but there is no graded frequency data available that takes object length into consideration as was the case in the English corpus study by Lohse, Hawkins & Wasow (2004). The frequency advantage for the PO order could be smaller for longer objects than for shorter ones in Norwegian, although the corpus study on English suggests the opposite with a near absence of the OP order for very long NPs. Memory-based processing accounts would predict higher reading times for the OP order, because of the higher cost associated with retrieving the verb from memory after a long object than a short one. If there is no retrieval of the verb at the particle and the non-local dependency is not established, we should see no effect of length. The SPR method and the comprehension questions that did not target the particle verb itself are not able to assess whether the dependency between the verb and the particle has been properly established in the OP condition or whether shallow processing occurred. A shallow parse that does not establish the particle as belonging to the verb and possibly misinterprets it as a preposition, i.e. the *på* 'on' in a sentence like *Marianne setter ringen på…* 'Marianne puts the ring on…" as the beginning of a prepositional phase like *…på bordet* '…on the table', would receive negative feedback for this interpretation at the following position in the spillover region that introduces the actual prepositional phrase. However, there was no evidence for reanalysis effects in the spillover region. Consulting the average reading times per segment in Table 7.9, we see that the spillover preposition even tends to be read faster in the OP condition compared to the PO condition with the exception of the short condition in L1. This suggests that the particle is correctly identified by the parser. Even if we assume shallow processing without the computation of the non-local dependency and the non-application of the EIC, the higher frequency of the particle-first order remains that renders the result of the overall analysis slightly puzzling. The results of the individual by-group analyses allow clearer conclusions.

The preference for the PO order in both length conditions in the L1 group is in line with the predictions of the EIC as this order allows faster assessment of the sentence structure or at least of the complex involving the verb and the object. It also reflects faster processing of the more frequent order. The length manipulation did not affect the processing of the particle. If L1 speakers predict a possible particle as soon as they encounter a verb that can take at least one particle, as was found for Dutch L1 speakers by Piai et al. (2013), encountering the particle after the verb should not be surprising for the Norwegian L1 group. Many particle verbs used in this experiment had some aspectual relation to the matrix verb, but for some the relation between particle verb and matrix verb was opaque. The retrieval of the matrix verb from memory and the reanalysis of the particle verb's content does not seem to be affected by the length of the object NP. Due to the ability of the L1 parser to predict an upcoming particle this reanalysis might be very swift, no matter how long the distance between the two elements. Models of parallel parsing could accommodate the parallel activation of all possible meanings of the verb with and without particle. The interpretation of no particle would be the one receiving most support initially, and the various interpretations with the particle would be subordinate to it. When the particle is reached, the parser could then boost the correct interpretation. The current design and method used in the experiment do not assess the presence of alternative verb interpretations, but a design similar to the one of the Piai et al. (2013) study or a cross-modal priming paradigm could shed more light on this question. As argued above for the overall result, memory based processing accounts would predict longer retrieval times for the verb in the long condition than in the short condition. The reading times of the particle, however, only showed generally slower reading times for the particle in the object-first condition no matter how many words intervened between verb and particle. The absence of a length effect in the L1 group could be due to type of length manipulation used. While corpus studies investigating length effects often use word, letter or syllable counts to measure length, e.g. measurement within the EIC is word-based, there are also authors that suggest that the distance between the two elements of a long-distance or non-local dependency should be estimated by the number of syntactic nodes that intervene between the two elements of a dependency. My length manipulation was based on word count, as the short condition involved one intervening word and the long condition involved four. However, the intervening material was always a noun that was modified with adjectives in the long condition. Keeping four words or 14 syllables in memory is not particularly challenging without additional side tasks, and could explain why I found no effect of length for the processing of the particle. In order to further investigate which type of measurement of syntactic complexity is more helpful in predicting elevated retrieval times, future research could compare, for example, adjectival NPs with relative clauses of the same length as intervening material. It seems that the preference for one order over the other might be highly item-specific and less dependent on pure counts of letters or syllables. Finally, the suggestions of descriptive Norwegian grammar were not reflected in the experimental data, as there were differences in reading times for the short condition for which descriptive grammar claims full interchangeability and there was no greater disadvantage for the OP order in the long condition compared to the short one, which is suggested by descriptive grammar.

The reading times of the L2 group show a null result with no effect of order on the overall reading times as both orders were read equally fast, with an identical increase of the reading times for the two orders in the long condition. The reading times for the particle were nearly identical across all four conditions reflecting no influence of either manipulation. Does this mean that German L2 processing of Norwegian is completely indifferent to the frequency difference between the two orders and to the processing advantages of the particle-first order? The offline rating data suggests that they are aware of a difference between the two orders and prefer the particle-first order. As already discussed for the L1 result, the present study does not allow to draw firm conclusions whether the L2 speakers finally formed the non-local dependency between the verb and particle, or if they interpret it as a preposition and do not attempt a reanalysis when the actual preposition is encountered. An initial interpretation of the particles as preposition would be in line with findings by Matlock & Heredia (2002) who compared sentences with the same phrase that could either be a particle verb (*John ate up the pizza*) or a verb followed by a preposition (*John ate up the street*). They found faster processing for the prepositional meaning than for the particle meaning in L2 speakers, whereas L1 speakers were faster with the particle verb. Evidence for a reanalysis in the spillover region is missing and the comprehension questions used in this experiment could also be answered correctly based on the meaning of the bare verb. Follow-up research focusing on the interpretation and syntactic integration of the particle is clearly needed. It is also worth noting that the object-first order as the only grammatical order in the

L1 of the L2 group might have actually helped the L2 participants in parsing the non-local dependency. Their experience with non-local dependencies involving particle verbs in their native language might have leveled out any processing disadvantages predicted for examples by the EIC. A reproduction of the experiment with another L2 group whose L1 either has no particle verbs at all or only allows the PO order would be needed to investigate this possible advantage due to L1 transfer.

Summing up, neither the interpretation of the overall analysis (effect of order only in the short condition) nor of the by-group analyses (general effect of order in L1, null result in L2) are in line with the recommendations of descriptive grammar. From a processing point, the two orders are not freely interchangeable in the short condition, nor does the processing of the OP order deteriorate with increasing length of the object.

7.5 Conclusion

In this study, native and non-native speakers exhibited a similar pattern in the acceptability rating task, but differed in the SPR task. The L1 group showed an effect of argument order in both tasks, while the L2 group only showed an argument order effect in their acceptability ratings. These results allow the following answers to the research questions.

Q3.1 Do native and non-native speakers have a general preference for one order of particle and object over the other?

a) Are the two orders interchangeable variations of each other for short NP objects (but not necessarily for longer ones) as is suggested by the reference grammar?

b) Is the verb – particle – object order the preferred order, as proposed by the EIC and Gries (2002) which state that it is easier to process?

c) If present, is this preference visible both in the online self-paced reading data and the offline acceptability rating?

There is a preference for the particle > object order, but this preference surfaced differently in the two groups. The L1 group showed a clear preference for the particle > object order with faster reading times and better acceptability ratings.

There was no evidence that the two orders are interchangeable variations of each other, as the rating difference was highly significant and also much bigger than the for example the difference in acceptability between the two German scrambling constructions investigated in Experiment 2a. However, the pilot study and Experiments 3a and 3b showed different results with regard to the influence of object length on the acceptability and processability of the OP order and do not allow a conclusive interpretation. The pilot study showed a strong association of object length and rating as the ratings for the OP order became less acceptable with increasing length. This result was found in both analyses using length as a categorical and as a continuous factor. The acceptability rating of Experiment 3a showed the opposite trend as OP orders with longer objects were rated more favorably than those with shorter objects in the categorical analysis. The reading time data of Experiment 3b showed no effect of object length on the reading times. While the general preference for the PO order is in line with accounts like the EIC and Gries (2002), these same accounts predict decreasing acceptability and processing speed with increasing object length. This latter association cannot be confirmed by the data gathered.

In contrast to native speaker performance, the L2 group only showed a general preference for the PO order in the acceptability rating task, but there was no significant difference in reading times between the two orders. The L2 group was also not sensitive to the object length manipulation in either of the two tasks.

Q3.2 Does the object – particle order get less acceptable for longer objects as suggested by the NRG, the EIC and also Gries (2002)?
a) Is this decrease in acceptability visible in online processing tasks in which a longer object might tax processing resources more than a short object…
b) …and/or is the decrease visible in offline rating tasks?

The larger-scale pilot study on the OP order involving 80 participants reported in section 7.1 had shown the predicted effects of object length with an overall decrease in acceptability for longer objects. However, the only effect of object length in the subsequent experiments was found in the acceptability rating task in the L1 group. This effect also went in the opposite direction of the predictions by the NRG, the EIC and Gries: acceptability actually increased for longer objects. In the self-paced reading task there was no evidence that longer objects caused the object-first order to be more taxing to online processing than shorter ones. Overall, the role of object length for particle placement remains inconclusive and requires further research.

Q3.3 Do L2 speakers of Norwegian with German as their native language show an advantage for the particle-final construction that is also present in their native language as was found in the production study by Bohnacker (2007)? Do they favor the object – particle order more than the Norwegian control group?

There was no overwhelming preference for the OP order in the acceptability ratings in the L2 group. Some participants showed extreme rating behaviors with regard to the 'German' order that were not found in the native group. These behaviors were either a complete rejection of the OP order or a stronger preference for the OP order over the PO order. The null result in the reading time data of the L2 group could have been influenced by their greater experience with OP orders as it is the only grammatical option in their L1. As there was only this one L2 group, any effects of transfer cannot be estimated conclusively and comparison with another L2 group from a different L1 background would be needed.

8 Study 4: Particle verbs in German

This chapter describes two experiments, an acceptability rating task and a self-paced reading task focusing on how L2 speakers of German process particle verbs in different syntactic contexts. The processing of particle verbs is a difficult task because their surface form depends on the syntactic context in which they appear. The learner has to associate two surface forms with the same lexical entry: *aufisst* and *isst auf* are both third person present tense forms of the infinitive *aufessen* 'eat up'. The first one is not split and occurs in non-V2 contexts, the second one is split and occurs in V2 contexts. Once the learner has learned this rule, he or she needs to be able to apply it to particle verbs, either in a verb-by-verb fashion or by generalizing the rule. Even if the verb is known as a particle verb, the learner still has to correctly identify the syntactic context in order to successfully apply the rule. In online processing the learner needs to have a sufficiently specified syntactic representation in order to assess whether a split is grammatically appropriate or not.

The acceptability rating task served the purpose of assessing the participants' knowledge about the splitability of the 24 particle verbs that were part of the self-paced reading task. Low performance on this task was used as an exclusion criterion for participants. If a participant does not know that a verb is a particle verb, then no effects of particle verb status on processing can be expected.

Particle verbs in general and German particle verbs in particular are well-discussed in the literature and there are some experimental studies regarding their processing, but to my knowledge there has not yet been a study looking into the processing of particle verbs by learners of German. Previous L2 studies on the processing of particle verbs often focused on lexical access and semantic differentiation, and investigated the difference between particle verbs and verbs followed by prepositions (Matlock & Heredia, 2002; Paulmann et al., 2015). These studies also mainly focused on English that only shows the alternation in particle placement investigated in the previous chapter. The present study uses the alternation in OV and VO word order in German to investigate how the changes in particle verb form caused by this alternation are processed by L2 speakers.

This chapter is structured as follows: Section 8.1 introduces the reader to the grammatical background of particle verbs in German and the Slavic languages spoken by the L2 group. Section 8.2 presents the acceptability rating task and its results that were used to assess the knowledge of the splitability rule for the investigated particle verbs. Section 8.3 then presents the self-paced reading task and Section 8.4 summarizes the findings of the two experiments. Finally, Section 8.5 provides the reader with an intermediate summary of part II of this thesis with the experiments on particle verbs in Norwegian and German.

8.1 Background: Particle placement in German

While verbs with particles and verbs with prefixes are not uncommon in the world's languages, the way German syntax handles particle verbs is a bit special. German particle verbs can occur in two basic forms: with the particle in a preverbal position (e.g. *aufessen*, eat up) or with the particle in a postverbal position (e.g. *isst auf*, eats up). The occurrence of the split is not random, but based on syntax as it has been linked to use of OV or VO order (see section 6.1). Non-topicalized particles appear in a fixed clause-final position. The separation of verb and particle that can be seen in (122a) has often been interpreted as resulting from particle stranding (S. Müller, 2002; Wurmbrand, 2000) within generative frameworks, or due to the application of a STAY constraint in Optimality Theory (Dehé, 2005). This clause-final position of the particle is also independent of the object's length. In non-V2 contexts with the verb in clause-final position, the particle always precedes the verb, but it can occur either with a conjugated main verb (e.g. in dependent sentences) (122b), as an infinitive (e.g. with modal verbs) (122c), or as a participle (e.g. in sentences with auxiliaries) (122d).

(122a) Martin **isst** das leckere, argentinische Steak, das ihm seine Mutter zum Geburtstag geschickt hat, damit er nicht so traurig ist, **auf**.
Martin eat the tasty Argentinean steak that him his mother to his birthday sent has, that he not so sad is up
'Martin eats the tasty Argentinean steak that his mother sent him for his birthday so that is not so sad up.'

(122b) Seine Mutter möchte, dass Martin das Steak **aufisst**.
his mother wants that Martin the steak up.eats.
'His mother wants that Martin eats up the steak.'

(122c) Martin will das Steak **aufessen**.
Martin wants the steak up.eat
'Martin wants to eat up the steak.'

(122d) Martin hat das Steak **aufgegessen**.
Martin has the steak up.eaten
'Martin has eaten up the steak.'

The participle of a particle verb is always constructed in the same way by placing the infix -*ge*-, used as a prefix in all other participles in German, between the particle and the usual participle form of the main verb.

Within the class of particle verbs there is one group of verbs that merits special attention. These are verbs that while having the same orthographic form behave either like a particle verb or a prefixed verb depending on the stress pattern. The verb *überziehen* can either be a particle verb behaving as in the examples (122a-d) above and mean 'to put on' when the stress is on the first syllable (*´überziehen*), or it is a prefixed verb meaning 'to overdraw' when the stress is on the third syllable (*über´ziehen*). Prefixes are never split from the verb stem and the participle also does not feature the -*ge*- infix.

The acquisition of German particle verbs poses several challenges to the L2 learner. The learner has to acquire the difference between particles and prefixes, which is comparatively easy as particles in German have the same form as prepositions and carry meaning while prefixes have no clear meaning by themselves (i.e. the prefixes *ent*- and *ver*- do not always add the same meaning to the verb stem). The learner also has to learn that particles can be split from the main verb and under which conditions this split occurs, namely in V2 contexts and for example not whenever the verb is conjugated. Acquiring the general V2-rule has been found to be difficult for L2 speakers, even for those from L1s that are also V2-languages (Bohnacker, 2006b; Rankin, 2014), which could further complicate the acquisition of German particle verbs. The splitability of German particle verbs also means that the learner has to be able to link two conjugated forms to the same infinitive (i.e. *isst auf* and *aufisst* both belong to the same infinitive *aufessen*). Finally, the learner also has to acquire the fact that the particle is always placed after the object, usually in sentence final position. This is considerably different from the acquisition task for English particle verbs. Here, L2 speakers have to learn that the same two words can either belong to a particle verb or a verb + preposition construction, that in the case of a particle verb the meaning of the original verb needs to be modified and that the particle can be placed either before or after a direct object.

Studies on the production of particle verbs by L2 speakers found that avoidance of particle verbs compared to simple verbs was modulated by proficiency and degree of similarity between the L1 and the L2 (Laufer & Eliasson, 2008): more proficient speakers and speakers from closely-related L1 backgrounds showed less avoidance (but see Siyanova & Schmitt, 2007 who found no effects of exposure in an L2 group with varied L1 backgrounds). The ERP processing study by Paulmann et al. (n.d.) found the same processing advantage for the figurative meaning of particle verbs over the literal meaning of verb + preposition constructions in English native speakers and Arabic learners of English suggesting that language distance might not play a role in this particular type of particle verb processing.

My choice of Slavic native speakers for the L2 group tries to reconcile the two findings. Slavic languages have an elaborate system of prefixes to modify the meaning of simple verbs in a similar way as particles are used in German. Prefixes are used to mark aspectual meaning, as the imperfective aspect is usually the bare infinitive (e.g. Russian *есть* (est') 'to eat'), whereas the perfective aspect often has a prepositional prefix added to this verb. By adding different prefixes, various meanings in the perfective aspect can be derived from the same verb (e.g. *пое́сть* (poest') 'to eat', *дое́сть* (doest') 'to finish eating', *съесть* (s'est') 'eat up'). Slavic prefixes are also used to mark directionality in verbs of movement which can take a large number of prefixes in both aspects. Some of the Slavic prefixes are also homonyms of prepositions, but unlike German particles they are never split from the main verb. The choice for Slavic speakers also tried to take into account the possible influence of L1-L2 similarity that has previously been found by some authors (Laufer & Eliasson, 2008).

The Slavic languages are not as closely related to German as, for example Dutch, and Slavic speakers are not familiar with the split component in V2 contexts. However, they are familiar with the principle of verb meaning modification by prefixes. This should put them at an advantage compared to speakers of Hebrew that were used in the study by Dagut & Laufer, (1985). L2 speakers of German with varied Slavic L1s are therefore likely to perceive the task as neither too easy (because of extreme similarities between L1 and L2) nor too hard (because of extreme differences).

Research questions:

Q4.1 Do L2 speakers of German with a Slavic L1 exhibit knowledge of the rule that particle verbs need to be split in V2 contexts? Is there evidence for a generalization of this rule, i.e. is the rule also applied to verbs the speakers are possibly unfamiliar with?

Q4.2 Do L1 and L2 speakers show online sensitivity to the incorrect splitting of particle verbs in non-V2 contexts, i.e. are reading times elevated in the ungrammatical compared to the grammatical conditions?

Q4.3 Does the syntactic context affect the processing of the ungrammatical split in non-V2 contexts, i.e. is there competition between the two possible finite verb forms (*aufisst* vs. *isst auf*) compared to contexts that only allow the finite form?

8.2 Experiment 4a: Acceptability rating task

This task was administered to check whether participants had correctly acquired the rule that particle verbs need to be split in a V2 context, and that the L2 speakers were able to identify 24 verbs used in both the acceptability rating task and the SPR task as particle verbs. The accuracy of the participants on this task also served as a measure for potential exclusion from the self-paced reading data, in the case of participants who misclassified too many verbs as non-splitable verbs.

Participants

The participants were the same as described under Section 5.2 with 39 participants in the L2 group and 33 participants in the L1 group.

Materials

The design was a simple design with split as a factor with the levels grammatical split and ungrammatical no split. There were 24 critical sentences in this task with 24 different particle verbs. Six different particles were used: *ab* (off), *an* (on), *aus* (off/out) and *ein* (in) occurred four times, *auf* (up/on) occurred five times and *weg* (away) occurred three times. Frequency counts were taken from the dlex database (dlexdb.de) and frequency of the infinitive was taken as both non-finite forms do exist, but only in the appropriate syntactic contexts.

The sentences were kept as simple as possible. All sentences were in simple present tense that provided the V2 context necessary for the split. 21 sentences had only an accusative object, the remaining three sentences had an accusative and a dative object. Subjects and objects were all nouns, and no proper names were used. A Latin square design was used and participants saw half the items with in the grammatically correct split condition (123a), the other half in grammatically incorrect no split condition (123b).

split condition
(123a) Der Chemiker nimmt die Schutzbrille ab.
the chemist takes the safety goggles off
'The chemist takes the safety goggles off.'

no split condition
(123b) *Der Chemiker abnimmt die Schutzbrille.
the chemist off.takes the safety goggles

The no split condition mimicked the way in which prefixed verbs are used in the Slavic L1s of the participants in the L2 group. The critical sentences from this experiment were mixed with the critical sentences of study 2 and fillers that were described in Section 5.2. The overall ratio between grammatical and ungrammatical sentences in the entire rating task was 24:26.

Procedure

As there was only one acceptability rating questionnaire with items from Experiment 2a and 4a, the procedure is the same as described under Section 5.2.

Predictions

The predictions for this task are based on whether the L2 group has acquired the rule that particle verbs need to be split in V2 contexts. There are no specific predictions for the L1 group as it can be assumed that all native speakers of Germans have acquired this syntactic rule.

A – If the L2 group has acquired the syntactic rule that particle verbs in German need to be split in V2 contexts, then they should perform like native speakers and rate the split condition as acceptable (=1) and the no split condition as unacceptable (=5).

B – If the L2 group simply transfers their L1 structure, then they would judge the no split condition as acceptable (=1) and the split condition as unacceptable (=5).

C – If the syntactic rule is acquired on a verb-by-verb basis, then L2 speakers might be more accurate in their judgments of more frequent verbs than of less frequent verbs.

Results

The complete dataset contained data from 33 participants in the L1 group and 39 participants in the L2 group that contributed altogether 1728 data points. Both groups showed the expected pattern that sentences in the split condition were rated as acceptable, while sentences in the no split condition were rated as unacceptable. As can be seen from Table 8.1, the difference between the L1 and L2 group in ratings and standard deviations is much smaller in the ungrammatical condition than in the grammatical condition.

	split condition	***no split condition**
L1 group (N=33)	1.15 (0.22)	4.32 (0.68)
L2 group (N=39)	1.37 (0.54)	4.43 (0.66)

Table 8.1 Average acceptability ratings for Experiment 4a, SDs in brackets

A between-groups ANOVA revealed no Group x Condition interaction ($F_1(1,70)=0.26$, $p=0.61$, $F_2(1,23)=1.52$, $p=0.23$), but main effects of both Group ($F_1(1,70)=3.92$, $p=0.052$, $F_2(1,23)=32.25$, $p<0.001$) and of Condition ($F_1(1,70)=927.74$, $p<0.001$, $F_2(1,23)=3357.96$, $p<0.001$). The main effect of Group is caused by the slightly higher average ratings by the L2 group. The main effect of Condition is clearly visible in table 46 above in the already mentioned better ratings for the split condition as compared to the no split condition. Post-hoc paired t-tests confirmed this result: $t_1(71)=-30.62$, $p<0.001$, $t_2(23)=-57.78$, $p<0.001$.

Post-hoc analyses

As the items of Experiment 4a had been combined with the items of Experiment 2a that required a more graded assessment of acceptability, the participants did not receive a binary task in the present assessment. However, the acceptability of the particle verbs runs more clearly along the lines of a binary grammatical/ungrammatical distinction than the acceptability of the scrambled structures used in Experiment 2a (see Poulsen, 2012). In order to assess response accuracy to see if any participants or items had shown particular difficulties, I recoded the reposes in a binary fashion. Responses of 1 or 2 were accepted as correct responses in the split condition and responses of 4 or 5 were accepted as correct responses in the no split condition. A response of 3 was coded as incorrect as it is in the middle point of the scale and therefore not particularly informative for the binary recoding. As there was not separate option to indicate a response of “I don’t know” or “I am unsure”, it is also possible that this midway option was chosen when the participant was in doubt. This recoding lead to the exclusion of 112 data points associated with an incorrect response and that were divided comparably among the L1 and the L2 group (54 and 58 answers respectively). The removal of responses affects the ungrammatical condition more than the grammatical one (46 data points for L1, 40 for L2).

	split condition	***no split condition**
L1 group	96.97% (0.06)	86.61% (0.28)
L2 group	90.81% (0.14)	86.32% (0.24)

Table 8.2 Accuracy scores by condition L1 vs. L2 group after binary recoding

As can be seen from Table 8.2, both participant groups were more accurate in the grammatical condition than in the ungrammatical condition. While both groups had basically the same accuracy score for ungrammatical items, there was a difference between the accuracy scores of the L1 and L2 groups of about 6% for grammatical items. An main effect of Condition was the only effect found in the overall ANOVA after a Bonferroni correction ($F_1(1,70)=5.34$, $p=0.024$, $F_2(1,23)=18.36$, $p<0.001$).

Participants also provided corrections for items that they had judged as unacceptable. The corrections provided by the participants again allow insight into problematic items and in how far ratings are related to the actual manipulation. While the number of uninformative “3” answers was comparable across groups, the L1 group had less errors than the L2 group in the remaining dataset (11 vs. 49).

Of the 11 errors in the L1 group, four occurred in the split condition and three were replacements of the particle for the same verb *anstellen* ‘to hire’, changing it to the synonymous particle verb *einstellen*. The remaining error in the split condition was the replacement of an entire particle verb with another regular verb. All seven errors in the no split condition stem from the same participant who did not use the entire rating scale by rating even highly ungrammatical filler sentences as comparatively acceptable and rating more than half of the sentences in the no split condition with “2”. This rating behavior had no match in the entire L1 group.

The L2 group made 49 errors that were equally distributed across the conditions: 25 in the split condition and 24 in the no split condition. The difference in accuracy scores seen in Table 8.2 is the result of the smaller number of available data points in the ungrammatical condition after the removal of mid-scale answers. Only 11 L2 participants performed without error, and the 49 errors were distributed over the remaining 28 participants. The majority of these 28 participants committed relatively few errors, answering at least 22 out of 24 sentences correctly. Four L2 participants accounted for 27 errors. Three of these participants had not provided any corrections to the rated sentences and generally showed a very mixed rating behavior without any clear pattern even for the ungrammatical filler sentences. The errors of the fourth participant all stemmed from the no split condition. The corrections that the participants had provided demonstrated clear knowledge of the split rule in V2 contexts as every item in the no split condition had been corrected to contain a split independent of the rating that had been given.

As some studies had found effects of proficiency on the processing of particle verbs (especially with regard to the opaque/transparent meaning distinction) and the L2 group had shown some variance with respect to their Goethe scores, AoA and time spent learning German and living in Germany, I ran a series of linear regression to assess the influence of these factors on the L2 rating accuracy. There was a marginal effect of Goethe score ($df=37$, $t=19.3$, p=0.06), suggesting a numerical trend towards higher accuracy with increasing proficiency as assessed by the Goethe test. The outliers in Figure 8.1 below also demonstrate that a good score on the Goethe test does not necessarily also indicate a good performance on the acceptability rating task.

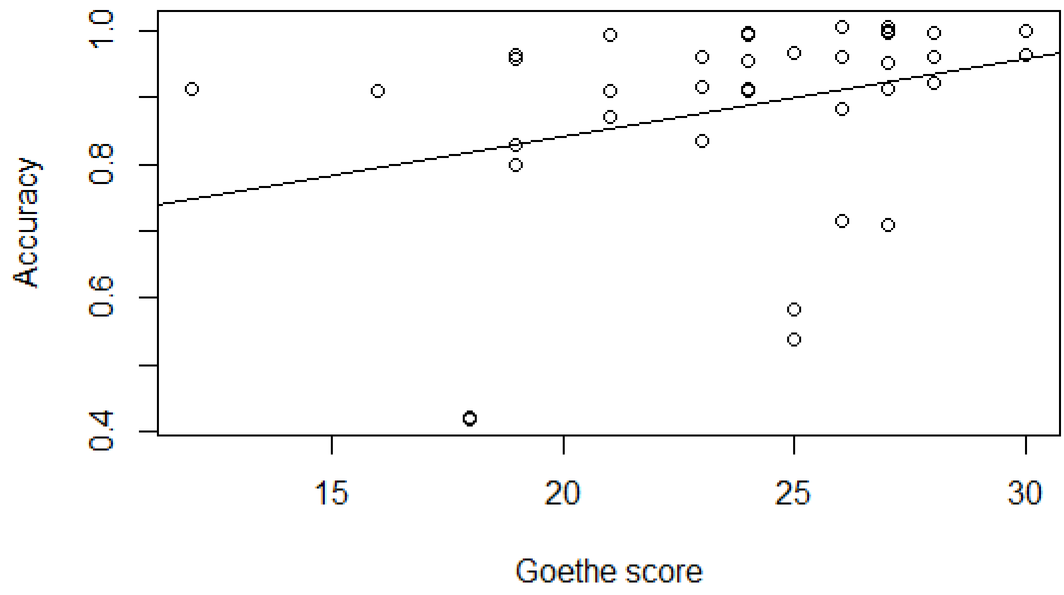

Figure 8.1 L2 accuracy scores plotted against Goethe score

Similar trends were found for analyses with AoA and years spent learning German. Time spent living in Germany had no influence at all on the accuracy scores. The results of all linear regressions can be found in Appendix B.

In order to investigate whether verb frequency played a role in the accuracy of the L2 group's responses, I ran a linear regression using the log-transformed lemma-frequency for each particle verb provided by the dlex database as a predictor for the accuracy of each verb across both conditions. The accuracy for individual items ranged from 71.8% to 100% and the linear regression showed a marginally significant effect of frequency (t=1.97, p=0.06) with higher accuracy scores for more frequent verbs (see Figure 8.2).

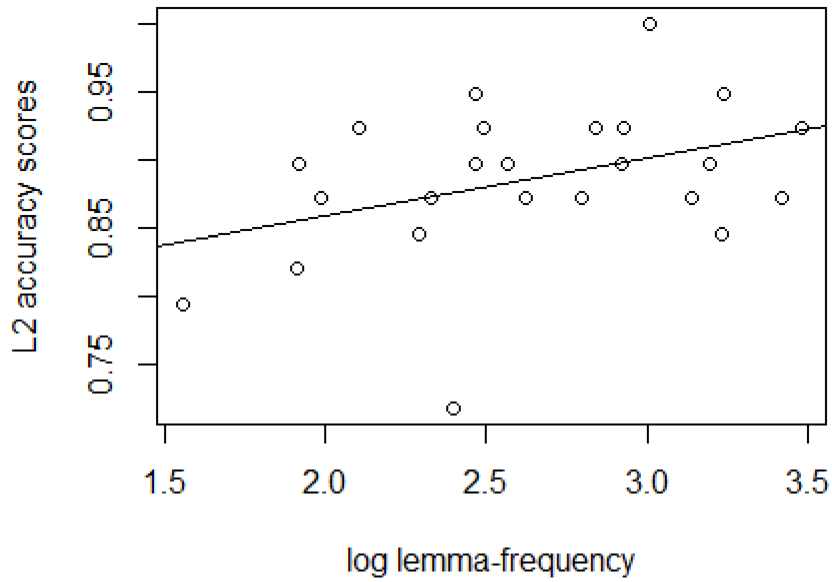

Figure 8.2 Accuracy scores by verb frequency for the L2 group

Discussion

The main aim of this task was to check whether all participants knew that the 24 verbs used in the experimental sentences were particle verbs that require a split in V2 contexts.

The native group performed as expected and correctly identified all verbs as particle verbs. The ungrammatical no split condition was correctly rated as less acceptable than the grammatical split condition, but as more acceptable than other ungrammatical filler sentences. A possible explanation could be that the conjugated, unsplit form does exist in embedded contexts and comprehension of the sentence is not hindered by the ungrammaticality. The ungrammatical fillers, however, contained manipulations that never appear in written language and would be indicators of errors and corrections in spoken language.

The majority of the L2 group also performed very well in this task as evidenced by accuracy scores of at least 86%. The successful completion of this task depended on the acquisition of the concept of a particle, especially for this type of L2 group whose native languages only have prefixed verbs, as well as the acquisition and application of the V2 rule which is a more generally applied rule of verb movement in German that extends far beyond particle verbs. A third of all

L2 participants performed flawlessly on this task and the vast majority of the rest only made occasional mistakes. This suggests no general problems with particle verbs or the V2 rule for the vast majority of speakers at the advanced proficiency levels like those that were tested for this experiment. However, the linear regression using lemma frequency as a predictor of accuracy scores showed that more frequent verbs were more often identified correctly by L2 speakers. Particle verb status is assumed to be part of a verb's lexical entry. If an L2 speaker has no lexical entry for a specific verb, particle verb status apparently cannot be deduced from its surface form by means of a share particle the speaker is already familiar with. Being familiar with verbs such as *aufsetzen* 'put on', *aufblasen* 'blow up' or *aufessen* 'eat up' did not seem to help in identifying the verb *aufsagen* 'recite' as a particle verb. This does not seem to happen in the L2 group as errors were spread across all verbs and not clustered by particle. This inability to generalize and instead having to access the particle verb information in the lexicon for each verb individually might therefore slow down the recognition process and affect reading times in the following self-paced reading study. L1 studies like Smolka et al. (2009) had found evidence for a connection between a base verb and all possible particle verbs derived from that base verb, independently of semantic relatedness in the mental lexicon. If L2 speakers cannot recognize that an unknown verb, e.g. *aufsagen*, consists of the particle *auf* plus the base verb *sagen* and is split into these parts under V2, then the representations of these two verbs (*aufsagen* and *sagen* 'say') in the mental lexicon should also not be related. This difference between L2 and L1 speakers forms one of the core problems of L2 acquisition research of particle verbs in general (see Alejo González, 2010). An interesting follow-up study to investigate the ability to generalize particle verb status could be a sentence completion task using nonce verbs with existing particles in V2 contexts. If native speakers use a general mechanism for particle splitting, they should also apply it to nonce verbs. If L2 speakers assign particle verb status individually to each verb and only after having gathered evidence for a split, they should struggle more with this task.

Three L2 participants had comparatively low accuracy scores and had misjudged up to 14 verbs. Their rating behavior was compatible with an influence of the L1 as they tended to rate sentences in the ungrammatical no split condition as acceptable and sentences in the split condition as unacceptable. These participants did not indicate that they were unfamiliar with the verbs in question, but it was clear that they did not know that these verbs were particle verbs, instead they treated them as prefixed verbs. The data of these three low-performing participants will be excluded from the self-paced reading data analysis as the experimental manipulation only has a possible effect if the participant knows that the verb is a particle verb. The data of the remaining participants that had been investigated more closely will be retained in the self-paced reading analysis as their 'faulty' judgments either reflected a systematic adjustment of the rating scale, or their corrections demonstrated clear knowledge of the split rule and its application to the verbs in question.

8.3 Experiment 4b: Self-paced reading task

This task measured the online processing of German particle verbs by L1 and L2 speakers. Unlike the offline acceptability rating task, it did not investigate particle split in V2 contexts, instead, it investigated the processing of ungrammatical splits in auxiliary and embedded sentences.

Participants

The participants were the same as described under Section 5.3. The same five participants that had been removed from the self-paced reading analysis due to low accuracy scores were also removed from this analysis. The data of three additional L2 participants was removed due to their low accuracy rates in the previously reported acceptability judgment task. The remaining group consisted of 31 L1 speakers and 33 L2 speakers.

Materials

The 24 particle verbs tested in the acceptability rating were also used for the items in the self-paced reading task. The self-paced reading task employed a 2x2 design. The first factor manipulated was split with two levels (split, no split) that were also used in the acceptability rating task. The second factor was syntactic context with the levels modal clause and embedded clause. As both are non-V2 contexts, in the self-paced reading task the no split condition is grammatical and the split condition is ungrammatical. Each trial was preceded by two contextualizing sentences, introducing the agent and the particle verb used in the critical sentence. The particle verb always occurred in the first of these two sentences in a V2 context and the particle was therefore split from the verb and in sentence-final position (see 124a).

The experimental sentences in the modal condition (124b+c) began with a proper name (half of them male, half of them female) which were chosen from a database of popular German baby names of the year 1986 close to when the majority of participants were born. Names were chosen in a way to avoid too much orthographic overlap or confusion between male and female names (e.g. Christian – Christiane). This proper name was then followed by an auxiliary and the object of the sentence. The object was followed by the particle verb, which was either a single word (no split condition) or two words (split condition). To move the critical region away from the end of the sentence, an adverbial phrase consisting of two or three words was added after the particle verb. Sentences in the modal condition contained eight or nine words in the split and no split conditions respectively. Due to differences in the length of the object and the final adverbial phrase, the critical region was unfortunately not always in the same position across experimental sentences.

The experimental sentences in the embedded condition (124d+e) contained the entire sentence of the modal conditions without the modal verb. The sentences began with an NP as the subject of the main clause, followed by the main verb and the connector *dass* ‘that’ introducing the embedded clause. Sentences in the embedded condition contained 11 or 12 words in the respective split conditions. For the same reasons as stated above regarding the modal condition, the critical region was not always in the same position across sentences.

Lead-in sentence

(124a) Sebastian **nimmt** jeden Monat fünf Kilo **ab**. Insgesamt sind es schon 25 Kilo, aber er ist immer noch sehr schwer.
Sebastian takes every month five kilograms off. Altogether are it already 25 kilograms, but he is always still very heavy.
(Sebastian loses five kilograms each month. Altogether it’s already 25 kilograms, but he is still very heavy.)

Modal, no split condition

(124b) Sebastian soll nochmal fünfzehn Kilo **abnehmen** diesen Sommer.
Sebastian shall again fifteen kilogram off.take this summer
(Sebastian should lose another fifteen kilograms this summer.)

Modal, split condition

(124c) *Sebastian soll nochmal fünfzehn Kilo **nehmen ab** diesen Sommer.
Sebastian shall again fifteen kilogram take off this summer

Embedded, no split condition

(124d) Die Ärztin verordnet, dass Sebastian nochmal fünfzehn Kilo **abnimmt** diesen Sommer.
The doctor prescribes that Sebastian again fifteen kilograms off.takes this summer
(The doctor prescribes, that Sebastian loses another fifteen kilograms this summer.)

Embedded, split condition

(124e) *Die Ärztin verordnet, dass Sebastian nochmal fünfzehn Kilo **nimmt ab** diesen Sommer.
The doctor prescribes that Sebastian again fifteen kilograms takes off this summer

In order to assess whether participants paid attention to the task and had understood the experimental sentences, they were asked 16 comprehension questions that queried the content of the experimental sentences. Half of them expected a negative answer. These questions either contained a synonym of the critical verb or a verb denoting the opposite action. The comprehension question for the above item was:

Comprehension question

(124f) Soll Sebastian fünfzehn Kilogramm zunehmen?
Shall Sebastian fifteen kilograms to.take?
(Should Sebastian gain fifteen kilograms?)

These comprehension questions were used instead of acceptability judgments in order to ensure that participants would continue to read all experimental sentences attentively for comprehension. Together with the items and questions from Experiment 2b, the fillers and the practice items, participants read 63 sentences and answered 25 comprehension questions. Altogether 20 sentences were ungrammatical: 12 experimental items all from experiment 4b, seven fillers and one practice item. After the practice session, participants were told that the experiments also contained ungrammatical sentences and that this was intentional. As the knowledge of the particle verbs was assessed in the acceptability judgment task, these verbs were not included in the vocabulary test described in Section 5.3. The randomization procedure and the fillers were the same as described in Section 5.3.

Procedure

The procedure for the overall experiment and the self-paced reading task were the same as described under Section 5.3.

Predictions

A – If participants are aware of the ungrammaticality of the particle verb split independent of the sentence type, then the reading times in the split condition should be elevated at the particle signaling the detection of the ungrammaticality.
B – If the availability of two surface forms for finite particle verbs (split and no split) causes a general competition between the two forms due to co-activation, then reading times for the embedded condition would be higher compared to the modal condition that requires the non-finite form for which there is only the unsplit surface form. Detection of the ungrammaticality would take longer in the embedded condition compared to the modal condition.
C – If the parser fails to detect the ungrammaticality of the split, then the particle should be read just as fast as an unrelated preposition in the spillover position in the grammatical no split condition.

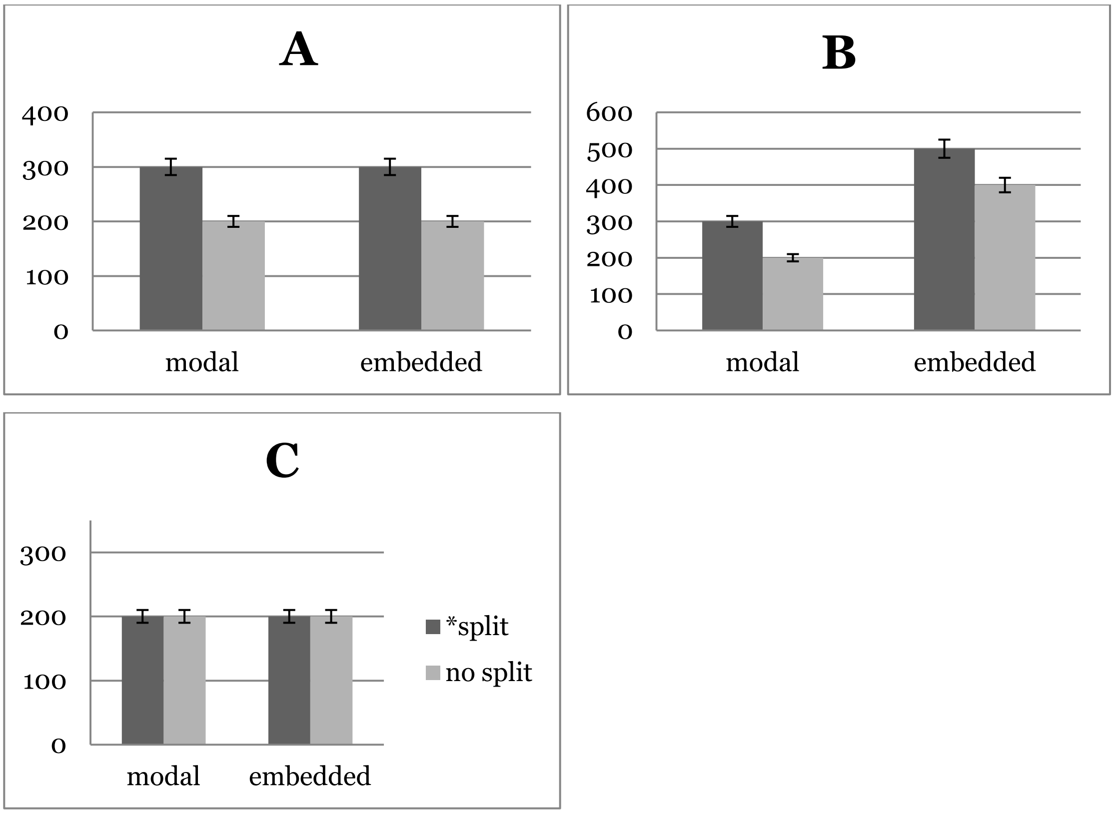

Figure 8.3 Idealized reading time patterns for the particle region (Experiment 4b)

Results

Comprehension questions

There had been 16 comprehension questions corresponding to questions after 2/3 of the experimental items, and 1152 data points for the whole group of participants tested (L1 = 33, L2 = 39). I repeat the presentation of response accuracies from Section 5.2 here. Accuracy scores for all participants were on average 75.78% (SD: 0.13). The L1 group reached 80.11% (SD: 0.11) accuracy and the L2 group 72.12% (SD: 0.14) accuracy. These averages were calculated including all participants. The averages in Table 8.3 reflect the recalculated averages after altogether eight participants had been excluded. Five participants had been deemed unsuitable due to low accuracy scores and likely lack of attention or task comprehension. Another three participants were removed from the analysis due to low performance in the acceptability rating task and likely problems with either the particle verbs used in this task or particle verbs per se (see Section 8.2).

	L1 group (N=31)	L2 group (N=33)	Both groups
Experiment 4b	81.05% (0.1) (range: 56.25 – 100%)	75% (0.12) (range: 56.25 – 93.75%)	77.93% (0.41)

Table 8.3 Mean accuracy scores for comprehension questions of experiment 4b, SDs in brackets

Accuracy scores of the remaining participants were above chance level and indicate that the L1 group was more accurate than the L2 group. This level of performance indicates sufficient comprehension of all experimental sentences. No additional participants were excluded from this set.

Data cleaning

After the removal of the abovementioned eight participants from the dataset, 1536 experimental sentences remained. The length of the experimental sentences varied by condition, and between eight and twelve data points were collected per sentence. The position of the critical region within the sentence also varied from item to item. Each individual participant's data was then adjusted to remove trials containing particle verbs that the participant had judged incorrectly in the acceptability judgment task. This resulted in the removal of 23 data points (1.5%), leaving a dataset of 1513 experimental trials. Extreme values and outliers were then removed for each segment separately. Due to the differences in length and position of the manipulation between the individual items and conditions, the data analyzed from every experimental sentence starts at four segments before the experimental manipulation. To account for the systematic difference in length between the verbs in the four conditions, - i.e. the fact that verbs in the split condition were always shorter than verbs in the no split condition, and verbs in the modal condition were always longer than verbs in the embedded condition - all analyses were run on residual reading times. In this method a linear regression is used to calculate predicted reading times based on the length and position of each individual word and actual reading times are then compared with these predicted values. Negative residual reading times indicate that reading was faster than predicted, positive values indicate slower than predicted reading. The residual reading times were calculated using all reading times with the exception of the practice items and running a linear regression with word length, position of the word within the sentence and subject as factors. Extreme values had been removed from the whole dataset by visual inspection before residualization.

Outliers among the residual reading times were identified based on a +/- 2.5 SD range around the participant mean per segment. A table with the cutoff points for extreme values and number of outliers removed for each segment can be found in Appendix B. The region of the matrix verb and the spillover region that contained either the particle in the split condition or a preposition in the no split conditions lost the most data as a result of this procedure, with more than 20 values deleted per segment already before the residualization and another 32 and 43 values deleted during the removal of outliers. The highest percentage of removed data for any segment was 4.49%.

Main analysis

The critical region in this experiment was the main verb and the following particle for split conditions and the following preposition for the no split conditions. As the position of the verb, however, varied among the experimental sentences, the overview over the reading times will be centered on the verb with previous and spillover segments. One item had only one segment following the particle verb, therefore the data analysis cannot go beyond this point as not all items would contribute data. The two or three regions preceding the matrix verb contained the same words and are therefore easily comparable. Any regions beyond this contained different words in the different conditions (e.g. name and auxiliary) complicating a direct comparison. Table 8.4 shows the residualized reading times for the core segments per condition per group. The region containing the split particle is shaded. For my main analysis I focused on the matrix verb and the following segment. The residualization process estimates reading times for each participant individually and eliminates any effects of generally slower reading times in the L2 group that were found in the previous three SPR experiments.

L1 group (N=31)						
auxiliary condition						
	Precritical 4	Precritical 3	Precritical 2	Precritical 1	Matrix verb	Spill/Particle
No split	28 (38)	-32 (54)	-38 (39)	-64 (42)	-74 (41)	32 (31)
*split	30 (41)	-24 (37)	-39 (32)	-47 (35)	-24 (36)	55 (49)
embedded condition						
	Precritical 4	Precritical 3	Precritical 2	Precritical 1	Matrix verb	Spill/Particle
No split	-13 (47)	-33 (45)	-37 (26)	-65 (35)	-70 (51)	26 (38)
*split	-17 (36)	-28 (28)	-37 (23)	-66 (33)	-30 (23)	42 (41)
L2 group (N=33)						
auxiliary condition						
	Precritical 4	Precritical 3	Precritical 2	Precritical 1	Matrix verb	Spill/Particle
No split	-53 (94)	-109 (81)	-85 (76)	-44 (91)	46 (145)	-4 (93)
*split	-49 (91)	-104 (70)	-83 (76)	-37 (88)	-1 (117)	-12 (103)
embedded condition						
	Precritical 4	Precritical 3	Precritical 2	Precritical 1	Matrix verb	Spill/Particle
No split	-97 (92)	-117 (89)	-84 (81)	-93 (69)	11 (121)	-19 (99)
*split	-104 (74)	-88 (63)	-88 (64)	-84 (80)	-16 (102)	-12 (79)

Table 8.4 Mean residual reading times across conditions for L1 and L2, SDs in brackets, darker shade indicates the particle

A graphical rendition of this table can be found in Appendix B. Between-groups ANOVAs run on the segments preceding the matrix verb showed main effects of Group for precritical regions 4, 3 and 2, reflecting more negative residual reading times in the L2 group than in the L1 group. There were also main effects of Type at precritical regions 4 and 1. At precritical region 4, both groups showed slower reading times for the auxiliary condition and at precritical region 1, the L2 group shows the same slower reading times for the auxiliary condition. This latter effect is absent in the L1 group, leading to a Group x Sentence Type interaction. It has to be noted that while the residualization process serves to even out differences in word length and word position within the sentence, it does not address differences in word type which most likely are responsible for the effects of sentence type in the precritical regions, as different types of words were compared, especially at the more distant precritical regions 4 and 3. A full overview of the ANOVAs conducted for all segments can also be found in Appendix B.

The between-groups ANOVA run on the residual reading times for the matrix verb showed a Group x Split interaction $F_1(1,62)=15.03$, $p<0.001$, $F_2(1,23)=15.73$, $p<0.001$ and a main effect of Group ($F_1(1,62)=13.94$, $p<0.001$, $F_2(1,23)=29.06$, $p<0.001$), all other effects were $F<1$. As there was no effect of sentence type, the following by-group analysis collapsed the data of the two sentence type conditions and only investigated the effect of the grammaticality manipulation in the two groups separately. Paired t-tests showed that the interaction found in the between-groups ANOVA is due significant effects of the grammaticality manipulation that go into opposite directions as can be seen from Figure 8.4. The L1 group has significantly faster reading times for the grammatical no split condition ($t_1(30)=52.25$, $p<0.001$, $t_2(23)=57.2$, $p<0.001$), whereas the L2 group has slower reading times in the grammatical no split condition, with a marginally significant result ($t_1(32)=3.94$, $p=0.056$, $t_2(23)=3.55$, $p=0.07$).

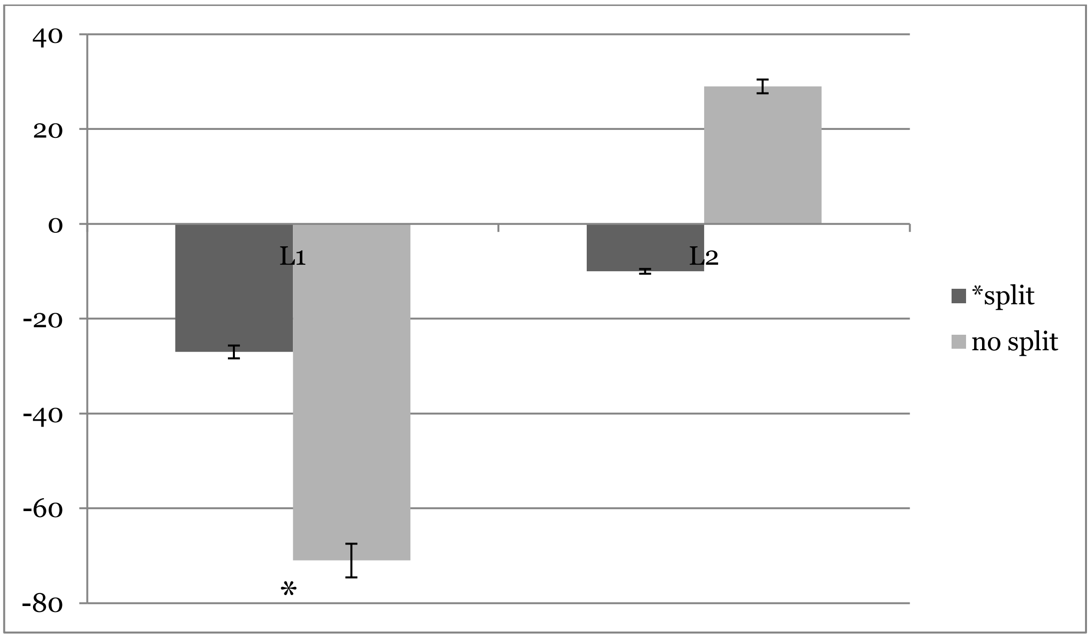

Figure 8.4 Residual reading times for the matrix verb

At the matrix verb the ungrammaticality of the split should only be apparent if the parser predicted the use of a particle verb. The ungrammaticality only becomes unambiguously apparent in the following segment that contrasts the split particle in the split condition with an unrelated preposition in the no split condition. Here, the by-groups comparison showed an effect of split in the L1 group ($t_1(30)=10.32$, $p=0.003$, $t_2(23)=3.8$, $p=0.06$) and no effect of split in the L2 group, both ts<1.

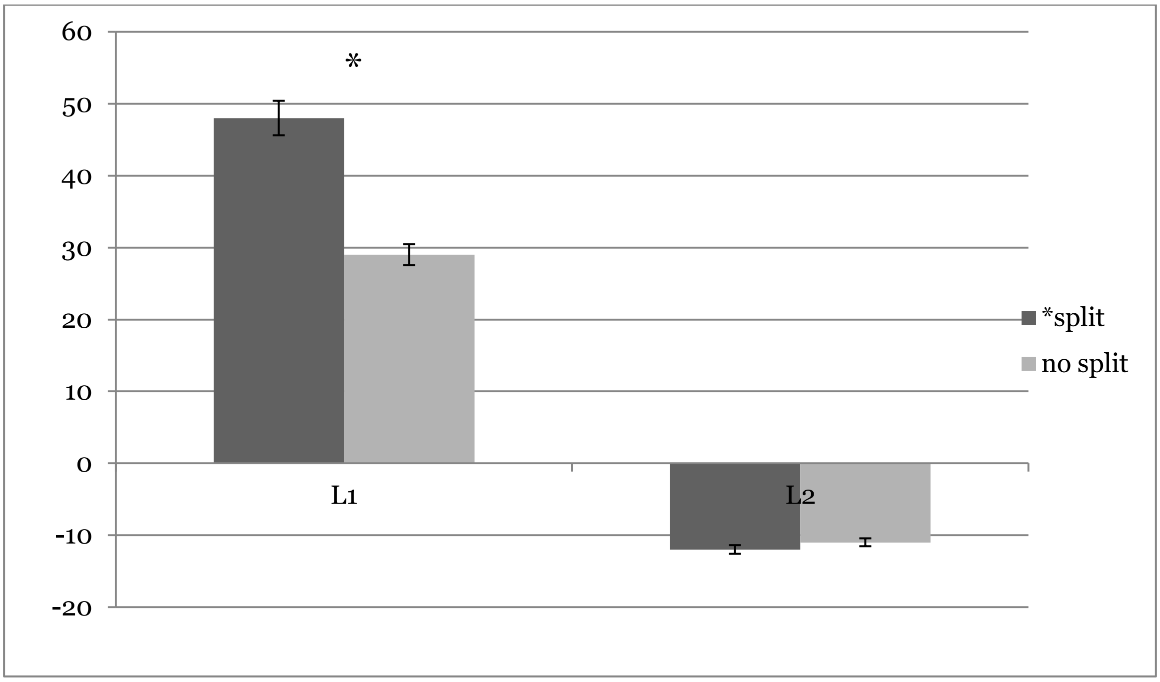

Figure 8.5 Residual reading times for the particle (*split condition) or the spillover preposition (no split condition)

A comparison of Figures 8.4 and 8.5 shows that in the L1 group residual reading times changes from negative values, i.e. faster reading than predicted, to positive values, i.e. slower reading times than predicted. The ungrammatical split condition always causes slower residual reading times. The residual reading times of the L2 group stay almost the same across both segments in the ungrammatical split condition and change from a positive to a negative value in the grammatical no split condition.

Post-hoc analysis

All of the L2 participants selected for the SPR task had shown to be proficient at identifying particle verbs and applying the V2-rule in the offline acceptability judgment task. Yet, the overall result of the L2 group's SPR data suggests an online insensitivity to the grammaticality of the critical sentences. As the participants exhibited great variation in their biographical data with regard to their AoA and overall exposure to German, I explored the influence of these factors on the online processing to investigate whether more experience with German could lead to a reversal of the pattern found in the L2 group and bring it closer to the native pattern. The results of the various linear regressions can be found in Appendix B. Effects of biographical variables were only found on reading times for the region of the verb. The biggest effect on reading times at this point was found for the variable "time spent living in Germany" (t=-2.84, p=0.008). A negative value in Figure 8.6 below indicated higher residual reading times in the ungrammatical condition which is the pattern found in the native group, and a positive value indicates longer residual reading times for the grammatical condition. A longer stay in Germany is no guarantee for a complete reversal of the effect found for the overall group, but the decrease in reading times for the grammatical no split condition gets smaller with increasing length of stay.

Figure 8.6 Reading time differences for the verb region plotted against years living in Germany (L2 data only)

There was no effect of subject-specific variables on the reading times for the particle/spillover region, and the reading time difference clustered more around a small difference of +/- 50 than in the preceding verb section.

Discussion

The self-paced reading data showed different processing behaviors for the native and the non-native group on the matrix verb and the following particle or spillover region. Neither group showed an effect of sentence type manipulation, suggesting that the processing of the particle verb in the finite condition (embedded sentences) was not affected negatively by the availability of two finite shapes for particle verbs (a split form and an unsplit form).

The results of the L1 group are in line with prediction A that assumed a general effect of grammaticality, but no reflection of the sentence type. The slower reading times for the ungrammatical condition at the particle indicate that participants did not expect a particle to follow after the matrix verb, as a particle in this position would be ungrammatical under all conditions. The significant difference between the particle in the split condition and the unrelated preposition in the no split condition suggest that the particle was not interpreted as a preposition, but that it was linked to the preceding matrix verb, creating an ungrammaticality. While the ungrammaticality could be unambiguously identified at the particle, the L1 group already showed an effect at the preceding matrix verb. The negative value of the residual reading times at the verb indicate that a verb was expected at this point, but the grammatical no split condition elicited faster reading times than the ungrammatical condition. In theory, a particle verb as well as a bare verb could have been possible at this point. However, L1 participants might have predicted the actual particle verb based on its appearance in the contextualizing sentence and the prediction of particles in native speakers found by Piai et al. (2013). If L1 speakers did indeed predict a particle verb at this point, they were able to detect the ungrammaticality before the encounter of the split particle which could explain the slower reading times for the ungrammatical condition at the verb and the following region.

Unlike the L1 group, the L2 group showed different reading time patterns at the two points of manipulation. At the verb, the L2 group did show an effect of grammaticality, but going in the opposite direction of prediction A. L2 participants were actually slower in reading the grammatical no split condition compared to the ungrammatical order. As the reading times had been residualized, the difference in length between grammatical and ungrammatical conditions – that is to say, the grammatical condition contained more letters than the ungrammatical one – cannot serve to explain this effect. L2 speakers might take more time to process the additional lexical information provided by the particle compared to the bare verb. Their residual reading times for the grammatical condition were also positive indicating that they took longer than expected to read the particle verb compared to the bare verb which yielded more or less the reading times predicted by residualization. Despite receiving the same contextualizing sentences as the L1 group, the L2 speakers did not seem to expect a particle verb at this point. Furthermore, residual reading times at the split particle were the same for the split and no split conditions, suggesting that participants processed particles and unrelated prepositions in the same way. One might argue that the faster reading times for the split condition are the result of syntactic priming caused by the appearance of the particle verb in its split form in the framing sentence. If this were the case, participants would have been primed to expect the particle to follow, causing faster reading times for the particle in the spillover region. Since this was not the case, it seems that L2 speakers were not sensitive to the grammaticality manipulation introduced by splitting the particle from the verb in a non-V2 context. L2 participants did not seem to identify the particle as a particle that needs to be connected to the preceding verb, causing a grammatical violation in the process. However, the elevated reading times for the unsplit particle verb compared with the bare verb suggest a sensitivity to the additional lexical content provided by the particle

The results of post-hoc explorations of subject-specific biographical variables such as AoA and length of exposure to German only showed an effect of these variables on the reading times for the verb section. The results were in line with exposure-based models of L2 processing: longer residence in German and an earlier AoA were beneficial to a faster processing of the grammatical verb. More experience with German seems to facilitate the processing of the additional lexical information provided by the particle. Whether it also facilitates the recognition of the ungrammaticality remains unclear, as none of the biographical factors had an influence on the processing of the split particle, which was generally processed at the same speed as the unrelated preposition.

8.4 Conclusion

The two experiments on German particle verbs yielded comparable results for the L1 group, who showed a sensitivity to the grammaticality manipulation in both cases. The L2 group, in contrast, showed different results in the offline and online tasks. Although they were able to correctly identify the experimental verbs as split verbs in the acceptability rating task, there was no evidence that they were sensitive to the grammaticality manipulation in the reading task.

Q4.1 Do L2 speakers of German with a Slavic L1 exhibit knowledge of the rule that particle verbs need to be split in V2 contexts? Is there evidence for a generalization of this rule, i.e. is the rule also applied to verbs the speakers are possibly unfamiliar with?

The majority of the L2 speakers was able to correctly identify particle verbs and to apply the split rule in V2 contexts. There was no evidence of a generalization of the split rule, and particle verb status seems to be assigned in a verb-by-verb fashion by L2 speakers as knowledge of another verb with the same particle was not used to identify an unknown verb as a particle verb.

Q4.2 Do L1 and L2 speakers show online sensitivity to the incorrect splitting of particle verbs in non-V2 contexts, i.e. are reading times elevated in the ungrammatical compared to the grammatical conditions?

L1 speakers showed clear evidence of a detection of the ungrammaticality reflected by elevated reading times for the ungrammatical condition both at the main verb and the particle/spillover region. L2 speakers did not show online sensitivity to the ungrammatical split in either syntactic context. On the contrary, they actually exhibited longer reading times for the grammatical particle verb over the ungrammatical bare verb that could be caused by the processing of additional lexical information. This effect was mitigated by increasing L2 language experience.

Q4.3 Does the syntactic context affect the processing of the ungrammatical split in non-V2 contexts, i.e. is there competition between the two possible finite verb forms (*aufisst* vs. *isst auf*) compared to contexts that only allow the finite form?

Neither the native nor the non-native group showed any influence of the morphological form. There was no slowdown in the embedded condition which used the finite verb that could possibly be associated with two shapes compared to the modal condition that requires the non-finite form for which there is only one possible shape. There was no evidence that the presence of the split finite form in the framing sentences and the lexical activation associated with it influenced the later processing of the non-split finite form.

8.5 Intermediate conclusion: Particle verbs

The two experiments on particle verbs reported in chapter 7 and 8 revealed similar results regarding the processing of particle verbs.

In both experiments, the L1 group showed sensitivity to the experimental manipulation in the offline and the online task, whether it was the order manipulation in the Norwegian experiment reported in Chapter 7 or the split manipulation in the German experiment reported in Chapter 8. In the Norwegian experiment there was a clear advantage for the particle > object order to be more acceptable and more easily processed. In the German experiment the ungrammaticality of the split condition was reflected in elevated reading times in the L1 group.

The patterns of the L2 groups are not as similar across the two sets of experiments as the patterns of the L1 groups. Both L2 groups performed more or less natively in the offline task. They were native-like in that they showed an overall higher acceptability for the particle > object in Study 3 and that they were able to apply the "split under V2"-rule for particle verbs in study 4. However, offline results were not native-like in that L2 participants showed no sensitivity for the length manipulation in Study 3 and did not seem to be able to generalize the V2-rule to unknown particle verbs in Study 4. The online data showed non-nativelike processing patterns throughout.

The results of Study 4 were clear in that the L2 group as a whole was not sensitive to the grammaticality manipulation in the way that the L1 group had been, and instead showed elevated reading times for the grammatical condition. The results of Study 3 were less clear. On the one hand, the results of the overall analysis did not allow a separate analysis for native and non-native speakers, suggesting that the patterns did not differ significantly. However, separate analyses for the L1 and L2 data were motivated by the higher variance and higher averages in the L2 data that had a bigger influence on the ANOVA than the L1 data. The separate analysis showed a null result in the L2 data and a clear preference for the PO order in the L1 group. Interactions with groups in ANOVAs usually do not turn out significant if one of the groups has a null result, as they test for opposite patterns in the data.

If we analyze the L2 data separately from the L1 data in both online studies, it is not clear whether L2 speakers actually form the dependency between verb and particle. This formation should have shown effects of order in Study 3, and should have triggered the detection of the ungrammatical split in Study 4. Neither effect was found in the L2 data, whereas it was clearly present in the L1 data.

Study 3 investigated optional word order variations that are open to personal preference, whereas study 4 investigated the application of a strict syntactic rule that is either applied correctly or not. The offline application of the syntactic rule and the preference was possible for the non-native speakers, but their application during online processing seems to be more difficult. This division into almost native-like behavior in offline tasks in the face of non-native-like behavior in online processing, whether it results from a different processing pattern or complete insensitivity to the manipulation, is found fairly commonly in L2 research (see e.g. Coughlin & Tremblay, 2012; Foucart & Frenck-Mestre, 2010; Patterson, Trompelt, & Felser, 2014), and is therefore not surprising.

The following general discussion aims to bring together the findings from all four experiments and seeks to fit them in with current L2 processing models. It also seeks to point out directions for further research and questions that could not be addressed in this thesis.

9 General discussion

The aim of this dissertation was to investigate and compare the processing of non-canonical word orders in Norwegian and German by native and non-native speakers. The results were mixed as the L2 groups showed native-like sensitivity to non-canonical sentence structures in only some experiments, whereas the L1 groups showed clear effects of non-canonicity across all experiments. The following discussion aims to bring together the results of all four studies and to put them in the context of L2 sentence processing theory, and, in those cases where it is possible, linguistic theory in general. I will begin by addressing the overarching research questions that were stated in the introduction of this thesis. I will then summarize the findings of all four studies conducted for this thesis and discuss their relevance for theories of native and non-native sentence processing. Finally, I will highlight individual aspects of the particular experimental manipulations used in my studies that warrant closer inspection, such as the role of animacy, frequency, and context. Section 9.1 provides an outlook on new research questions that arose from the present findings and suggestions how future research can investigate these questions. Section 9.2 contains the final conclusion of my thesis with the main theses I developed from my data.

Q1 Are non-canonical word orders more difficult to process than canonical ones?

a) Does this processing difficulty surface for all types of word order variation investigated (topicalizations, scrambling, particle verbs)?

b) Is there a difference between native (L1) and non-native (L2) processing of word order variations?

c) Does the non-native parser detect all order manipulations equally well? Does it use the same cues as the native parser to identify non-canonical orders?

Native speakers showed effects of processing difficulty associated with reanalysis of non-canonical word orders across all types of orders investigated. Non-native speakers only showed signs of reanalysis in the study on object topicalizations.

General discussion

There was no reading time difference between canonical and non-canonical orders for L2 speakers in the studies on scrambling or particle verbs. In the study on object topicalizations, non-native speakers used the cues main verb/NP2 order and NP animacy in the same way as the native speakers. However, animacy did not facilitate the processing of scrambled orders. Object case marking and particle verb status were also used differently by native and non-native speakers, that is to say that non-native speakers made reduced or even no use of these cues. The implications of this difference in cue use will be discussed in more detail below.

Q2 Are non-canonical word orders perceived and rated as less acceptable than canonical ones?

a) Is the difference in acceptability the same for all types of word order variation investigated or are some more comparable in acceptability than others?

b) Can native and non-native speakers identify gradient acceptability equally well or is a bimodal distinction of grammatical/ungrammatical easier to identify?

Similarly to the online processing data, native speakers perceived all non-canonical orders as less acceptable than the corresponding canonical ones. The difference in acceptability was more pronounced for the particle shift in Study 3 than the scrambled objects in Study 2. The non-native speakers showed a similar trend as they did differentiate the canonical and non-canonical order to a similar degree as the native speakers for the particle shift. However, there was no significant difference between the two scrambled orders in Study 2 in which also the native speakers had made a much smaller difference. Gradient acceptability seems to be more difficult to detect for non-native speakers than the bimodal distinction of grammaticality in Study 4 which the non-native group handled as well as the native group. However, the successful distinction in Study 3 suggests that a more nuanced perception of acceptability can be possible also for non-native speakers.

Q3 Are patterns or preferences that emerge in the offline tasks also reflected in the online processing behavior?

a) Does native-like performance of the L2 group in the offline task equal native-like performance in online processing?

b) Does the ability to achieve a native-like pattern depend on the type of word order variation?

Native speakers' offline and online performance showed identical patterns in those studies in which the two investigated the same structures. Lower accuracy scores and less acceptable ratings corresponded to slower reading times. This was not always the case for the non-native group. Offline and online performance showed the same patterns in Study 1 and Study 2. However, in Study 1 this identical result was a sensitivity, whereas in Study 2 the non-native speakers were equally insensitive to the manipulation in both tasks. In Study 3, the patterns of offline and online task differed as non-native participants showed a clear preference in the offline task, but not processing advantage in the online task. Native-like performance in both tasks was achieved for the order variation that causes the biggest changes to sentence structure and content, i.e. object topicalization. For the other structures investigated that caused more subtle changes to sentence structure or content, i.e. scrambling and particle placement, the non-native group was only partly able to perform like the native group. They correctly identified the non-canonical order as acceptable, but did not show sensitivity to changes of acceptability related to object length or information structure. It seems that the chance to achieve native-likeness depends on, among other things like proficiency and probably individual cognitive differences, the structure investigated. This aspect of non-native performance will also be discussed more deeply below.

Summary of experimental findings

In Study 1 on object topicalization in Norwegian, the OVS order was challenging for native and non-native speakers alike. Both groups had similar error rates in the agent identification task and their reading time patterns in the experimental OVS conditions showed a strong influence of animacy. Reanalysis was completed faster if the sentence-mediate subject was animate than if it was inanimate. An error analysis in the agent identification task showed a stronger influence of animacy in the L1 group than in the L2 group. Some L1 participants might have transferred the contexts in which object topicalization is allowed in ambiguous structures (only when the object is lower in animacy than the subject) to the unambiguous structures investigated. No such trend was visible in the L2 group.

Study 2 on object scrambling in the German midfield showed a small, but robust preference for the DAT > ACC order in the acceptability rating task and also a reading time advantage for this order at the main verb in the self-paced reading task in the L1 group. The L2 group showed no significant difference between the two orders neither in the acceptability rating nor in the self-paced reading task. Both orders were rated as highly acceptable.

Study 3 on Norwegian particle shift revealed a general preference for the particle > object order in the L1 and L2 group. This preference was additionally modulated by object length in the L1 group. The by-group analysis of the SPR data showed a general processing advantage for the particle > object order that was independent of object length in the L1 group, and no RT difference in the L2 group. The overall analysis only showed a reading time advantage in the short object condition and no difference for long objects.

The grammaticality manipulation of Study 4 on German particle verbs was detected by L1 and L2 groups alike in the offline task involving V2 contexts. L2 participants were able to correctly identify the verbs used as particle verbs, and knew the rule that they had to be split in V2 contexts. The same particle verbs did not trigger an effect of ungrammaticality in the online processing task in the L2 group. The L1 group, in contrast, showed a clear effect of ungrammaticality with longer reading times for the split particle compared to an unrelated preposition. The L2 group had shown a slowdown in reading times on the verb itself with slower reading times for the particle verb compared to the bare verb, likely caused by the additional lexical content that had to be processed. Table 9.1 summarizes the results of the L1 groups for all eight experiments.

		Norwegian	German
Non-canonical objects	**Judgment task**	OVS higher error rate than SVO	DAT > ACC more acceptable than ACC > DAT
	SPR	OVS _inanimate with slowest RT	DAT > ACC read faster than ACC > DAT
Particle verbs	**Judgment task**	particle > object more acceptable than object > particle	unsplit particle verbs were rated as unacceptable
	SPR	particle > object read faster than object > particle[15]	grammatical particle verbs were read faster than ungrammatical particle verbs

Table 9.1 Summary of experimental results for the L1 groups only

Implications for theories of (native) sentence processing

As already stated in Section 2.1, which reviewed processing models, not all experiments in this thesis contribute equally well to the evaluation of different types of processing models. I will go through the five types of processing models that were introduced in the review and use my experimental findings to evaluate them.

Syntax-first models (e.g. Bornkessel & Schlesewsky, 2006a; Frazier & Fodor, 1978), which assume a dominance of syntax in online processing and expect longer reading times whenever a reanalysis of the syntactic structure is needed, are able to explain the majority of the L1 findings. The OVS order in Study 1 requires a reanalysis of the syntactic structure as exemplified in Figures 4.1 and 4.2, and an additional thematic reanalysis. This reanalysis was found in the L1 and the L2 data. The ACC > DAT order of Study 2 requires a reanalysis, possibly even at the points. First at the dative object when the sentence structure has to be revised from a transitive to a ditransitive sentence, or at the main verb when then thematic roles are assigned and the non-canonical ordering of the objects becomes evident. The latter kind of analysis was found in the L1 data. The discontinuous order of verb and particle in Study 3and Study 4 should cause longer reading times as they require a reanalysis of the verb as an actual particle verb and the following establishment of a filler-gap dependency in the syntactic representation. In Study 4, this process additionally entails the detection of the ungrammaticality. Evidence for this kind of reanalysis was again found in the L1

[15] This was true for both length conditions only in the separate by-group analysis that was not motivated by a Group x Order interaction. In the overall analysis there was no difference in reading times for the long condition.

data. The absence of a reanalysis effect in much of the L2 data could be explained by a lack of depth or specification in the syntactic representation. If, for example, the particle is not identified as belonging to the verb, no reanalysis and filler-gap dependency creation takes place, hindering the identification of the ungrammaticality and rendering the distinction between the two orders in Study 3 void. This would be in line with approaches to L2 processing that assume weaker syntactic representations and less use of a syntactic processing route in L2 speakers, like the SSH (Clahsen & Felser, 2006c). However, the strong influence of NP animacy on reanalysis that was found in Study 1 does not fit well with strict syntax-first models. It can be accommodated by weaker versions of this approach or by models assuming parsing that is lexically-driven (e.g. MacDonald, Pearlmutter, & Seidenberg, 1994).

Interactionist or constraint-based models assume the integration of several sources of information during online processing, among them syntax, the lexicon and construction frequency. The predictions of the Competition Model (MacWhinney, 2005) regarding the use of the cues word order and NP animacy in Study 1 were fully borne out: more competition among the two cues led to increasing reading times in L1 and L2 speakers alike. Constraint-based models also predict a processing advantage for the DAT > ACC order in Study 2, as it follows the Dative constraint, the hierarchy of thematic proto roles as suggested by Primus (1994), or an independent Animacy constraint as suggested by Tomlin (1986). However, this advantage was only found in the L1 data. While L2 speakers might not be sensitive enough to morphological case marking to be affected by a violation of the Dative constraint, they should be able to use the lexical information of animacy, as was found in Study 1 and other L2 processing studies (e.g. Jackson & Roberts, 2010). From a lexical point of view, processing a particle verb could be more effortful compared to the bare verb as it contains additional lexical information, or in the case of particle verbs that have an opaque relation with the bare verb, a complete change of the semantic content. A greater processing effort of this type was found in the L2 data of Study 4. The L1 data had shown faster reading times for the particle verb as the bare verb was ungrammatical at this point. This distinction could again be interpreted along the lines of the SSH with more lexically-driven parsing in L2 and more syntax-driven parsing in L1. Construction frequency is often central to models of this type and it makes a clear prediction – more frequent orders are easier to process than less frequent ones. One hallmark feature of non-canonical word orders is their low frequency compared to canonical orders. In fact, for many of the non-canonical orders used in this thesis their frequency in corpora is so low that especially L2 speakers are unlikely to have encountered them very often (Kempen & Harbusch, 2004a). In light of this difference in frequency, it is difficult to explain why L2 speakers showed no effects of frequency in their acceptability ratings for German midfield scrambling and no reading time differences between canonical and non-canonical orders in Study 2 and 3. In a similar argument as above, it could be claimed that, in order to be able to detect the frequency difference, L2 participants need have sufficient syntactic knowledge of the construction. Again, if case marking is not identified by the parser, the difference in frequency between the two scrambled orders based on NP animacy is perhaps less striking than when case marking can be used properly. In the experiments in this thesis, frequency was always confounded with canonicity, but some experiments that used frequency as a variable found that canonicity exerted a bigger influence on processing than frequency (Bornkessel et al., 2002; but see Kempen & Harbusch, 2003). Interactionist models seem fit to explain the multitude of factors (frequency, animacy, case, object length etc.) that have been found to interact in the processing (and production) of order variations by a number of studies (e.g. Jäschke & Plag, 2016; Keller, 2000; Pappert et al., 2007; Stoops et al., 2014), but especially the lack of frequency effects found in the L2 data is problematic. Equally problematic is data that shows a frequency-acceptability gap (Kempen & Harbusch, 2005) as processing difficulty has been found to affect acceptability judgments (Fanselow & Frisch, 2006) and orders with a low frequency should be more challenging to process, causing a bigger difference also in acceptability.

The two memory-based processing accounts reviewed (Gibson, 2000; Lewis & Vasishth, 2005) both assume that sentences containing displaced elements are harder to parse, because the displaced element has to be kept in memory and reactivated at its original site. The processing difficulty is either estimated based on item similarity or distance between the element and its integration site. In the first case, multiple similar constituents, e.g. NPs, compete for reactivation and choosing the correct item is more difficult than in cases in which constituents are dissimilar. In the latter case, an increasing distance between a filler and its gap site puts a strain on working memory once the parser has to go back to reactivate the filler. The similarity-based account is interesting when comparing the results of Study 1 and Study 2. In both studies, sentences contained two NPs that required correct thematic role assignment to mark a sentence's canonicity, and that differed regarding their animacy. In Study 1, these NPs did not even bear morphological case marking, rendering them more similar than the object NPs used in Study 2 that had different unambiguous case marking. However, the SPR data of Study 1 showed a difference between the OVS orders, and the OVS_animate condition differed only little from the SVO conditions. Comprehension accuracy was not a dependent variable, and it is therefore not possible to estimate effects of NP similarity on the repetition of the sentence during the answer process, as (Rösler et al., 1998) suggested a silent repetition of the target sentence during the question process. If NP similarity was affected by animacy, this should also be reflected in the accuracy data when used as a proper dependent measure. If NP similarity only concerned NP surface form, the two OVS orders should not differ because the retrieval process of the topicalized object NP is identical in both cases. The reading time difference found in the L1 data in Study 2 fits a similarity-based processing account, if we assume that case marking is sufficient to clearly distinguish the two object NPs in the working memory of the L1 speaker. At the main verb, when thematic roles are assigned, all three NPs in the sentence are clearly distinguished, canonicity is identified, and the filler-gap dependency is established in the case of the non-canonical order. Using ambiguous case marking would lead to greater similarity and greater processing difficulty. Insufficient use of case marking could also be the source of the null result in the L2 group. The NPs are not sufficiently distinguished in working memory, causing high interference during thematic role assignment, and the non-canonical structure cannot be identified. With the experiment as it was used in this thesis, it is not possible to verify this claim as the comprehension questions did not question the final thematic role assignment nor was there a 'control' condition, using for example a mix of NPs and pronouns to make the constituents more dissimilar in working memory. When considering similarity-based processing accounts and the results of Studies 1 and 2, it is important to consider how non-canonicity was signaled in the two studies. In Study 1 the cue was the order of main verb and NP2, two highly dissimilar constituents, whereas in Study 2 it was the order of the two object NPs. Identification of the non-canonicity could therefore be harder in Study 2, as the two scrambled orders are overall more similar in their surface representation than the SVO and OVS orders in Study 1. The big difference in accuracy scores for the agent identification task of Study 1, compared with the small difference in the L1 group and lack of a difference in the L2 group for the acceptability rating task of Study 2, are in line with this prediction. The clear result in the acceptability rating of Study 3 that preferred the order without a filler-gap dependency and also used two dissimilar constituents (a particle and an NP) to signal the order difference also support the idea that a higher similarity in surface order makes the identification of a non-canonical order harder, or at least decreases the perceived difference between the two orders.

The distance-based processing account has more problems to accommodate the findings of this thesis. An increased distance between a filler and its gap should make the retrieval of the filler harder and result in more processing effort. In Study 1, the distance between the displaced object and its gap site are identical for both OVS orders. Yet, the OVS_animate condition showed very little additional processing effort compared to the SVO conditions, but a significant difference from the OVS_inanimate condition. In Study 2, the distance between the displaced accusative object and its integration site in the ACC > DAT condition was only two words (the determiner and the noun of the dative object) and this short distance was enough to result in a processing difference at the main verb. In contrast, the distance of eight words between the displaced inanimate object and its integration site in the OVS_animate condition of Study 1, covering a relative clause, the auxiliary, the subject NP and the main verb, did not result in a great reanalysis, although far more information had to be kept in working memory. Similarly in Study 3, the increase in object length from one word to four words did not affect the reading times for the particle in the particle-final condition in the separate L1 data analysis, although the distance between the verb and the particle increased and with it also the length of the non-local dependency. The analysis of the overall data of this study is even less compatible with a distance-based processing account as the advantage of the local dependency was only present in the short condition that should actually be less challenging for integration, as the distance is very short with only one intervening word and immediate integration is possible. In the long condition, there was no difference between the two orders in the overall analysis which is counterintuitive from a distance-based processing point of view. Study 4 did not feature a length manipulation, as the separated particle followed directly after the verb and allowed immediate integration. This immediate integration seemed to allow an easy detection of the ungrammaticality in the L1 group, but not in the L2 group.

Good-enough processing accounts of L1 processing (Ferreira et al., 2002; Ferreira, Engelhardt, & Jones, 2009) were described in response to conspicuous findings that indicated insufficient depth of syntactic representations when making grammaticality judgments or answering comprehension questions. During good-enough processing, speakers rely primarily on superficial information such as world knowledge, plausibility (i.e. a dog is more likely to bite a man than vice versa) and lexical information to make their parsing decisions and syntax is only used later to evaluate the fit of the parse and readjust it in case of counterevidence. The accuracy data of the agent identification task of Study 1 suggest that at least some participants in both participant groups used underspecified syntactic representations when choosing the agent. All sentences were globally unambiguous regarding their syntactic structure, but sentences with two logically possible agent NPs showed low accuracy scores in both groups, and sentences with an inanimate agent additionally had low accuracy scores in the L1 group. It is not possible to draw firm conclusions regarding good-enough processing in the SPR task of this experiment as there were too few comprehension questions, but the low accuracy scores that was found for comprehension questions from Study 1 (65%) suggest that they were more problematic than the questions from Study 3, which were part of the same experimental session and had an overall accuracy score of 93%. Good-enough processing can only be detected when the parser misjudges the content of a sentence, in order variations such as scrambling or the particle verb alternation, in which the content is not affected by the change there is no immediate punishment for good-enough processing. The difference in animacy between the two objects in the scrambling experiment is sufficient to assign the thematic role and clarify the action, and it is irrelevant for sentence comprehension whether the accusative object precedes or follows the dative object. The same is true for the order of particle and object in Study 3. Nevertheless, the L1 groups in both studies showed clear preferences for one order over the other, preferring the order with the higher frequency and lower syntactic complexity in both offline and online tasks, suggesting that good-enough processing did not take place. The grammaticality manipulation of Study 4 is more difficult to assess in the light of good-enough processing. On the one hand, whether a particle verb is correctly or incorrectly split does not affect the comprehension of the sentence and would allow good-enough processing. On the other hand, the discontinuous placement of the particle and the verb introduces the particle as a separate element that needs to be connected to something in order not to be superfluous within the sentence, especially in the ungrammatical split condition used in the SPR task. Good-enough processing in this case would mean that the particle is first superficially interpreted as a preposition and not reanalyzed as a particle that needs to be connected to the verb when the actual preposition is encountered, meaning the ungrammatical split remains undetected, but the sentence should nevertheless be perceived as overall ungrammatical as the bare verb did not fit the sentence and the doubled preposition was also ungrammatical. As the SPR task did not include a grammaticality judgment and used comprehension questions instead, it is not clear whether the L2 group that had not shown a reading time difference at the split particle, actually perceived the sentence as ungrammatical. The L1 group had shown a clear slow down in processing at the ungrammatically split particle, suggesting that they did not attempt good-enough processing and instead tried to integrate the particle immediately with the verb to compute a full syntactic representation. The findings of this thesis regarding L1 processing do not rule out the possibility of good-enough processing. It is, however, surprising that in situations in which the parser could get away with a good-enough parse without causing misunderstandings in communication, i.e. scrambling and particle shift, there were clear differences between the two variations in acceptability and processing behavior. Good-enough processing has been claimed to be a response to the demands of sentence processing during interaction in order to alleviate demands on the processing system. However, the main evidence for good-enough processing in my data comes from the agent identification task, an offline task without any time pressure. If good-enough processing is a strategy that is actively chosen by the processing system, I would expect the parser to make use of this processing strategy especially in cases in which there are no negative consequences for it. The data of Studies 2 and 3 do not fit with this suggestion. Good-enough processing could therefore rather be a signal of processing breakdown under pressure instead of an actively chosen strategy. The data of the L2 group, in contrast, shows many signs of good-enough or shallow processing, such as a lack of distinction between the order variations in Studies 2 and 3 and no sensitivity to the grammaticality manipulation in the Study 4. This is in line with the Shallow Structure Hypothesis (Clahsen & Felser, 2006a) and other theories of L2 processing that suggest an underuse of syntactic information in L2 sentence processing.

Implications for theories of L2 processing

Table 9.2 compares the results of L1 and L2 groups in the eight experiments. In three experiments, the results of L1 and L2 group were identical: The reading time patterns and accuracy scores in Study 1 did not differ, and the L2 group was largely able to identify particle verbs and the need for a split in V2 contexts. In another three experiments, the results of L1 and L2 differed: there was always a clear effect of the order manipulation in the L1 data but not in the L2 data. The two experiments of Study 3 on Norwegian particle verbs showed mixed results depending on the analysis chosen. It can also be seen from Table 9.2 that the L2 group's behavior was consistent across offline and online task only for the studies on non-canonical objects. Offline and online behavior of the L2 group did not converge in the studies on particle verbs.

		Norwegian	German
Non-canonical objects	Judgment task	L1 = L2	L1 ≠ L2
	SPR	L1 = L2	L1 ≠ L2
Particle verbs	Judgment task	L1 = L2 // L1 ≠ L2[16]	L1 = L2
	SPR	L1 ≠ L2 (L1 = L2)[17]	L1 ≠ L2

Table 9.2 Summary of experimental results comparing native and non-native performance

Nativelike processing in non-native speakers seems to be rare, and I will use those results that showed non-nativelike behavior to argue against the Fundamental Identity Hypothesis as proposed by Hopp (2007) and other theories

[16] The general order preference was the same, but the length effect was only present in the L1 group.

[17] When considering the results of the overall ANOVA that showed no interaction of Group and Order, there is no difference between L1 and L2 speakers.

that assume identical representations in L1 and L2 speakers. Hopp and similar accounts explain differences in behavior between L1 and L2 speakers with L1 transfer or performance factors, i.e. processing an L2 requires more resources than processing the L1, and L2 speakers reach the limit of the processing resources at an earlier point. When looking at the online processing data gathered for this thesis, we find that L2 speakers showed generally slower reading times than the L1 group across all four SPR tasks, suggesting that L2 processing is indeed slower and more effortful than L1 processing. However, when comparing the processing patterns, the one study in which there is no difference in overall pattern is the study on Norwegian object topicalization, which involves the longest filler-gap dependency of any of the four studies, and requires a syntactic as well as a thematic reanalysis for successful parsing. The reanalysis process and the creation of the filler-gap dependency in this construction should be more taxing to working memory and overall processing resources than the reanalysis processes in the other three studies that contain less complex syntactic structures.

The SPR tasks in the other three experiments showed null results for the L2 group. There was no reading time difference between DAT > ACC and ACC > DAT orders, nor between particle > object and object > particle orders. Meaning that L2 speakers did not show effects of factors such as the higher frequency of one order or the earlier resolution of the filler-gap dependency in the particle > object order that should make one order more easily processable. If L2 speakers only have more limited processing resources, but the same syntactic representations as L1 speakers, they should favor those order variations that put less strain on the parser. My data does not provide evidence for this theory. It could be argued that the null result in Study 3 is due to positive transfer of the German particle-final order and greater familiarity of the parser with this structure, cancelling the general processing advantage and higher frequency of the particle-first order. But without a second L2 group from a different L1 background that does not have this possible source of positive transfer this is highly speculative. The same argument is true when considering transfer of L1 information structure preferences in the scrambling study. If Slavic L1 speakers are more sensitive to the given > new ordering principle, they might prefer the ACC > DAT order, again making up for the frequency advantage and the canonical structure of the DAT > ACC. This question is harder to resolve as context was kept constant for both conditions, but did not strongly bias the ACC > DAT interpretation, nor did it lead to an actual reading time advantage for it. More research is needed to evaluate claims of transfer and disentangle possible effects of context in this particular case. In Study 4, L2 participants showed a null result on the split particle and the opposite effect of the L1 group for the preceding verb. The L1 group had apparently predicted a particle verb and showed faster reading times when this prediction was met, while the L2 group showed slower reading times for the particle verb, likely associated with the processing of additional lexical content. In the following region, the L2 group did not identify the ungrammatical particle and apparently did not attempt to connect the particle with the preceding verb, unlike the L1 group. The online data suggest that nativelike performance is possible even for a highly complex structure that should be more straining for processing resources than other simpler structures, going against the assumptions of the FIH. However, the results that showed non-nativelike performance were null results that cannot be interpreted convincingly and much could be blamed on transfer, an argument that cannot be refuted given the absence of a second L2 group. Why were L2 speakers seemingly unable to use frequency information and other ordering principles that have been claimed to hold crosslinguistically like the EIC (Hawkins, 1994)? I suggest that this is due to an underspecified L2 grammar and underuse of syntactic information during processing. If, for example, the parser fails to correctly identify the object case marking in the study on scrambling, it cannot apply a possible Dative constraint or frequency information as both structures receive a shallow Subject – Object1 – Object2 representation. Previous studies found that L2 speakers of German are able to use case marking (e.g.Gerth et al., 2015; Hopp, 2009; Jackson, 2008), however these studies compared subject and object case marking that could be a more salient distinction for L2 speakers than distinguishing the case marking of two objects. Previous studies on scrambling in the midfield often used judgment data and not online processing data of L2 speakers (e.g. Hopp, 2005). The same argument applies to the processing of particle verbs: If the parser fails to identify the non-local dependency between the verb and the particle, frequency information or the preference for a shorter dependency cannot influence the processing outcome. Native speaker data by Piai et al. (2013) on Dutch particle verbs suggested an automatic prediction of a particle whenever a verb could possibly take at least one particle. This experiment has not been conducted with L2 speakers yet and it is unclear whether they would show the same kind of prediction. If not, we could be dealing with two completely different mechanisms in particle verb processing: a forward-searching predictive mechanism in L1 processing, and a backward-searching mechanism that only considers a particle verb interpretation after the encounter of the particle in L2 processing. More research is needed to investigate this possibility that would also go against the FIH.

How can the nativelike processing in Study 1 be explained, if shallow processing with less use of syntactic information is assumed along the lines of the Shallow Structure Hypothesis (Clahsen & Felser, 2006a)? I suggest that it is the type of information that makes the difference. While the object topicalization of Study 1 requires a deep syntactic representation to correctly identify the non-canonical order, the difference between the SVO and OVS order is signaled by a change in surface word order of the main verb and the NP2. This change can also be detected in a shallow parse, whereas the scrambling study depends on the correct interpretation of the morphological case marking. The nativelike pattern in L2 speakers therefore does not contradict the SSH. Non-native speakers might always compute a shallow, syntactically underspecified parse first, but they receive compelling counterevidence for their parse that triggers reanalysis, whether this reanalysis results in a correct syntactic representation is not clear as comprehension accuracy was not a dependent measure. As already argued above in the section on good-enough processing in L1, the remaining three structures might not provide enough counterevidence to start a reanalysis process in L2 speakers, leaving the sentence's representation underspecified. This underspecification, however, has no consequences for comprehension. In this interpretation, taking away agreement morphology in Study 1 did not actually make the task harder for the L2 group, although they were deprived of powerful cues to sentence interpretation in their German L1. Instead, it seems they could focus on surface order information that is more easily detectable.

The offline data showed identical patterns for L1 and L2 speakers in two tasks. L2 speakers had the same problems with agent identification in OVS sentences as the L1 speakers, and they were equally able to apply the split rule to German particle verbs in V2 contexts. In the remaining two experiments, the L2 results showed similar trends as the L1 results. In the study on object scrambling, both orders were considered equally acceptable compared to a small, but reliable preference for the DAT > ACC order in the L1 group. This judgment does not constitute an incorrect assessment of the two orders, but it seems that L2 speakers are less sensitive to information structure orderings involving the Dative constraint. Hopp's (2009) study on midfield scrambling of direct objects across subjects had shown that near-native L2 German speakers were sensitive to the differences in information structure between the orders and that they used morphological case marking for their judgments. However, this study contrasted subject and object case marking, i.e. a violation of the Nominative constraint, and the two arguments additionally differed in animacy. An earlier study by Hopp (2005) on more complex scrambling of objects showed prolonged difficulty regarding the semantic and information structure motivations of syntactic reordering. The Dative constraint has been found to be a weaker constraint in native processing (Keller, 2000) and it might be too weak to be considered in L2 processing. Similarly, in the acceptability rating of Norwegian particle placement, the L2 group showed a general preference for the particle-first order, like the L1 group, but was not sensitive to the length manipulation. Based on assumptions of the EIC principle (Hawkins, 1994) and suggestions by the Norwegian Reference Grammar (Faarlund et al., 1997), acceptability of the particle-final order should decrease the more material intervenes between verb and particle. The EIC principle was developed to be applied crosslinguistically and based on general principles of processing economy. Given that L2 processing requires more processing resources, minimizing the processing effort should be in the interest of the L2 parser and there is no intuitive reason why L2 speakers should be insensitive to the EIC principle. When comparing the three acceptability judgment tasks and the performance of the L2 groups, the ability of the L2 group to perform in a nativelike way might depend on their definition of acceptability and grammaticality, a distinction that has been debated more generally in theoretical linguistics (see Poulsen, 2012). Despite being labeled an acceptability judgment task, the assessment of German particle verbs in Study 4 investigated the successful application of a syntactic rule, yet there was no range in acceptability as it is a clear question of grammaticality. A sentence containing an unsplit particle verb in a V2 context is grammatically incorrect and should receive a less acceptable rating than a sentence containing a split particle verb in the same context. Whether this rating then turns out to be 3, 4 or 5 on a 5-point-scale depends on individual rating behavior as some participants are strict raters, while others are more generous and even the repetition of the rating for 24 verbs can affect the absolute rating (see Zervakis & Mazuka, 2013). The acceptability ratings of Study 3 and 4 contained structures that were always grammatical and the order of their constituents is influenced by a myriad of factors such as NP animacy, order frequency, constituent length, discourse context, sentence intonation. Native speakers in my thesis and also in other experiments were found to be sensitive to these many factors in the production and processing of order variations (e.g. Kallestinova, 2007; Keller, 2000; Pappert et al., 2007), whereas non-native speakers seem to be less sensitive to these factors (Hopp, 2005; Jäschke & Plag, 2016; Park, 2011). Because of the many factors involved in making an acceptability judgment, finding a small difference in acceptability between two grammatical orders is likely harder for L2 speakers than a bimodal grammaticality distinction. L2 speakers have also been found to be more affected by task variables such as the presence of time pressure in judgment tasks compared to L1 speakers (Godfroid et al., 2015). The easier identification of a grammatical violation could also be due to how languages are taught in the classroom. A lot of attention is given to errors in grammaticality, probably because this type of error is highly perceptible and marks the speaker as an L2 speaker. Using a less acceptable order variation does not constitute an error in grammaticality. It is not 'bad German' or 'bad Norwegian', but it is less idiomatic and also less native-like. The high number of factors that influence constituent order in these cases further complicates the teaching of the appropriate usage of various similar orders. The investigation of methods to possibly increase awareness of gradient acceptability in L2 speakers is a task for applied linguistics.

The results of the agent identification task of Study 1 draw attention to another potentially important area in L2 research that could not be addressed in this thesis. The fact that the L2 group was able to perform like the L1 group in this task was not due to their exceptionally accurate completion of this task, but rather due to the fact that the L1 group also struggled with this task. Both groups showed similar amounts of errors and the variance found in accuracy scores was also comparable. One third of all participants had no problems with using the main verb/NP2 order cue to identify a topicalized object and performed flawlessly. The remaining two thirds showed a tendency to base their agent assignment more on NP animacy. Especially the lower performing participants of the L1 group seemed to apply topicalization preferences for ambiguous sentences that are based on animacy also to unambiguous sentences. Individual differences within the participant groups, independent of L1 or L2 status, might have a relevant influence on the performance in this task that needs to be further explored.

The role of context for the processing of non-canonical word orders

Topicalizations and scrambling are responses to demands of discourse and information structure, and occur on the syntax-discourse interface that is central to the Interface Hypothesis (Sorace, 2007; Sorace & Serratrice, 2009). However, context was not an experimental variable in this thesis, although the two studies on German provided context sentences during the SPR task. A supportive context has been found to improve comprehension accuracy of object topicalizations in L1 speakers (Kristensen et al., 2014), but the context used in Study 2 and 4 might have actually been more damaging than helpful as it might have biased the participant's reading behavior. The context in Study 2 introduced the accusative object, changing the information structure requirements and making it difficult to interpret the elevated reading times for the dative object that could reflect genuine problems with the dative, or just be response to a novel noun. The context in Study 4 introduced the split form of the particle verb that was to follow in the experimental sentence. This was done to bias participants towards this exact particle verb, but could have caused syntactic priming. As context was not manipulated, I will not use my data to evaluate claims that phenomena at the syntax-discourse interface are harder to acquire than others. The intricacies of object scrambling in the German midfield that reflect small changes in focus structure seemed indeed more problematic to the L2 speakers than object topicalizations which involve a bigger change in focus structure. But this could also be due to other reasons that are not related to their status as syntax-discourse interface phenomena, such as the already discussed underspecification of morphological case marking in the L2 group. Similarly, particle shift is not a phenomenon of core syntax, but is associated with influences of intonation and, relatedly, object length, at least in offline rating data. Are phenomena at the syntax-phonology interface also harder to acquire for L2 speakers in the same way as those at the syntax-discourse interface as suggested by the Interface Hypothesis? More research that actively manipulates the corresponding interface partner, i.e. discourse or intonation, is needed to evaluate the claims of the Interface Hypothesis.

9.1 Future directions

The findings of this thesis were able to shed some light on the processing of non-canonical word orders by L2 speakers. It seems that L2 speakers can show native-like judgments and processing for some non-canonical orders, but not for others. I will first address general questions that arise from the results of all the studies reported in this thesis and general suggestions for future research. I will then move on to more specific questions and suggestions for the individual studies.

The insensitivity to order changes in Study 2 on German midfield scrambling compared to the native-like performance in Study 1 on Norwegian object topicalizations warrants further investigation regarding the source of the insensitivity. Do L2 speakers struggle more to identify the non-canonical word order if it involves two objects and not a subject and an object? Related research on the English dative alternation showed that some L2 speakers did differentiate the two object orders (Jäschke & Plag, 2016; Marefat, 2005; Park, 2011). Interestingly, the Russian L2 group investigated by De Cuypere et al. (2014) did not differentiate the two orders of the English dative alternation. However, it has to be kept in mind that the English dative alternation differs from German midfield scrambling. The English alternation contrasts one order with two NPs with a second order with one NP and a PP, whereas the German order alternations involve two NPs in both cases. The studies on the English dative alternation also did not investigate online processing, but were mainly acceptability judgment studies and production studies. It is also not clear what role case marking plays. Again, in studies contrasting subject and object, L2 speakers have been found to be able to use case marking as a disambiguation cue (Gerth et al., 2015; Jackson, 2007), but studies on object alternations in L2 have either used languages without morphological case marking (English) or contrasted orders that differed on more than just case marking (Baten & De Cuypere, 2014). Overall, the insensitivity to canonicity manipulations in object scrambling and the role object marking in L2 processing plays warrant further investigation.

In a similar vein, the general role of the material that signals the canonicity manipulation is still unclear. The two studies in which the non-canonical order was signaled by constituents from different word classes, i.e. verb and noun in Study 1 and noun and particle in Study 3, showed effects of non-canonicity at least in the offline tasks. Study 2 in which non-canonical order was signaled by two nouns showed no sensitivity to the order manipulation. This question is closely related to the preceding one regarding the role of type of manipulation and type of disambiguation. If the constituents that signal non-canonicity are very dissimilar, surface order is sufficient to tell the orders apart. If however, the constituents are very similar, as is the case in German object scrambling, additional information, in this case morphological case marking, is needed to identify an actual change in the word order. In this thesis, the Norwegian experiments always used dissimilar elements, whereas the German studies used similar elements. A contrastive study within one language, using one group of participants, could shed more light on the role of constituent similarity in L2 processing of non-canonical word orders. This question is also connected to similarity-based models of sentence processing (Lewis & Vasishth, 2005) that assume more difficult retrieval of elements from memory, if several highly similar elements have been stored in memory, and models assuming shallow or good enough processing (Clahsen & Felser, 2006b; Ferreira & Patson, 2007) as non-canonical orders involving similar elements require more thorough parsing than those involving dissimilar items.

Three of the four studies in this thesis investigated grammatical word order alternations, whereas the remaining fourth study on German particle verbs investigated an alternation that was ungrammatical, at least in the syntactic context presented. The offline data of Study 4 and also the L2 judgment data by Hopp (2005) suggest that L2 speakers can correctly distinguish grammatical from ungrammatical non-canonical orders, despite their low frequency, high number of movements involved or, in the case of German particle verbs, the ungrammaticality of the surface form depending on the syntactic structure, i.e. the unsplit form of a particle verb is grammatically incorrect in V2 contexts, but correct in non-V2 contexts. However, the SPR data suggest that the L2 group was not able to identify the ungrammaticality during online processing. For L1 processing, Meng & Bader (2000b) suggested that the saliency of the ungrammaticality determines whether reanalysis is attempted at all or whether the parser abandons the analysis and judges the sentence as ungrammatical. It is not clear whether this can also be said about L2 processing and future research on ungrammatical non-canonical orders could shed more light under which circumstances the L2 parser is able to identify the ungrammaticality.

The following study-specific suggestions are intended to address questions that were left open by the results of the studies as they were conducted for this thesis. They mainly address matters of study design, experimental method, and participant groups.

Study 1

Study 1 on object topicalization in Norwegian showed a clear influence of NP animacy on the reanalysis process. Previous studies found that sentences with two animate NPs caused stronger garden-path effects than sentences with only one animate NP. Study 1 found an additional difference between sentences with only one animate NP, as animate subjects had a facilitative effect on reanalysis. Adding a condition with two animate NPs would make the SPR task more comparable to previous studies and at the same time evaluate the influence of competition between two suitable NPs on reanalysis. While sentences with only one animate NP have clear preferences regarding agency assignment, i.e. the animate NP is the preferred agent, sentences with two animate NPs are less clear and agency assignment could be delayed or reanalysis could be weaker compared to the conditions used in Study 1 of this thesis.

Introducing animacy as a within-item variable instead of a between-item variable in the agent identification task would allow to better study its role in offline agency assignment. The results of Experiment 1a suggested some difference between native and non-native speakers regarding their offline use of animacy, but the number of critical items was too small to draw any firm conclusions.

An elaboration of the agent identification task along the lines of previous studies within the CM framework (e.g. MacWhinney et al., 1984; Staron & Kail, 2004) with the introduction of globally ambiguous sentences and a more systematic variation of possible cues in Norwegian, such as animacy and definiteness, would allow the evaluation of the tentative cue hierarchy suggested in this thesis. It would also allow to compare cue use in globally ambiguous sentences of the NP1 V NP2 type investigated by Øvrelid (2004) with the unambiguous sentences used in this thesis. A study that considered individual differences between speakers, for example regarding working memory or syntactic integration ability as used by Hopp (2015), could shed some light on the high variation found especially among the native speakers in the offline task, and the possibility that some speakers transfer the contexts in which object topicalization is allowed from ambiguous to unambiguous contexts.

The comprehension questions of the SPR task in their form used in this thesis were clearly challenging for all participants. In order to maximize the information gained from the questions and make the task more participant-friendly, a picture verification task could be used instead that avoids a structural bias or complicated passive structures. Introducing this task after each trial would also provide more information about which trials were misparsed, and would allow the comparison of correctly and incorrectly answered trials.

Study 2

The cause of the null result in the L2 group in acceptability rating task as well as the SPR task remains unclear. One possibility for the SPR task is the context that was provided, but not manipulated for each condition. The presence of subject and accusative object in the context might have biased the L2 group more strongly towards the accusative-first order than the L1 group. As context was not systematically manipulated, the source of the elevated reading times for the dative object also remain unclear. Properly manipulating context using a match-mismatch design, and possibly a no context condition as baseline, could disentangle these effects and facilitate the assessment of L2 speakers' ability to apply the given > new and the dative > accusative ordering constraints.

Rerunning the experiment using only Bulgarian or Macedonian native speakers in the L2 group would be interesting from a transfer theory perspective. As in German, the object ordering preferences in Bulgarian and Macedonian depend on the form of the object (full NP vs. pronominal NP), but are the exact opposites of the order preferences in German. An additional manipulation of object type could reveal whether processing preferences develop for nouns and pronouns in the same way.

A task investigating the ability of the L2 group to correctly use case would be a very helpful addition to this experimental setup, as they did not show any reading time differences on the case-marked determiners and it is unclear whether the group could actually use the case information provided by the determiner. The correct identification of object case is crucial in the identification of the non-canonical order in scrambling.

Study 3

Object length had shown some effect on the acceptability ratings in the L1 group, but it did not influence their online processing. If the object length used in the SPR task was simply too short to cause measurable processing difficulty, introducing longer objects could be a solution. If the shifted material was not syntactically complex enough to cause processing difficulty, the comparison of a simple object NP with, for example a complex phrase with the same number of syllables might yield more revealing results.

Testing a second L2 group with an L1 background that has no particle verbs (e.g. Slavic speakers) would allow us to compare their processing results to the null result of the German L2 group, which may have been influenced by an advantage for the particle-final structure and L1 transfer.

Study 4

The acceptability rating task, as it was used in this thesis, was successful in assessing whether L2 speakers knew the verbs under investigation were particle verbs and that they needed to be split in V2 contexts. It did not, however, check whether they also knew that particle verbs do not need to be split in non-V2 contexts as they were used in the SPR task. A fill-in-the-blank task testing the correct use of particle verbs in various syntactic contexts might be more fitting. Introducing some nonce verbs to this task would allow assessment of ability of L2 speakers to generalize the split rule and apply it to unknown verbs.

Investigating particle verbs in self-paced reading is problematic, as it compares one word (no split conditions) with two words (split conditions). This can partly be addressed by either calculating residual reading times as done in this thesis or by using phrase-by-phrase presentation. Phrase-by-phrase presentation could also make the ungrammaticality more salient.

9.2 Conclusion

The eight experiments reported in this thesis provide evidence that non-canonical word orders pose different challenges for native and non-native speakers. Native speakers showed clear differences between canonical and non-canonical orders in all four studies. Non-canonical orders caused longer reading times, lower accuracy scores and were perceived as less acceptable by native speakers. Non-native speakers did not show this general effect of non-canonicity, as only some non-canonical orders caused disruption in sentence processing, whereas others did not differ from a canonical order in terms of processing speed or acceptability. This leads me to the following theses:

Not all non-canonical word orders are identifiable for L2 speakers.

Non-canonical orders that involve major changes to the content of the sentence and that contrast two salient thematic roles (e.g. subject and object) can be identified by L2 speakers as found in Study 1 of this thesis and other previous studies on subject/object ambiguities (Gerth et al., 2015; Jackson, 2007; Jackson & Roberts, 2010). This type of non-canonical order causes a similar processing slowdown in L1 and L2 speakers. In contrast, non-canonical orders that do not change the overall content of the sentence, e.g. via thematic role reassignment, but rather correspond to small changes in information structure as in Study 2 or Study 3, are harder to identify for L2 speakers. They may have learned that both orders are grammatical as seen in the acceptability rating task of Study 2 or in the results of Hopp (2005). They may even develop a preference for the canonical order as seen in the acceptability rating task of Study 3. However, during online processing both orders are treated as equivalents. I suggest that this is due to a shallow syntactic representation that is not specified enough to differentiate canonical and non-canonical orders. In the case of midfield scrambling as in Study 2, the thematic roles of the object arguments remain underspecified, resulting in two seemingly optional orders of non-subject NPs.

Whether non-canonical word orders involving small changes to sentence content result in a difference in acceptability in L2 speakers depends on how they are disambiguated.

Order variations that involve only small differences in content, like particle shift or midfield scrambling, are clearly differentiated by native speakers with a preference for the canonical order. This preference may be numerically small, as was the case in Study 2, but it is reliable.

L2 speakers are able to develop a preference for the canonical order to nearly the same extent as native speakers, as long as the difference between the two orders is signaled by salient information that can also be identified during shallow parsing. A word order change that involves two constituents from different word classes, e.g. particle and object NP as in Study 3, is more easily detectable for L2 speakers than a change that involves two NPs that need to be differentiated based on case marking as in Study 2. This thesis is in line with findings underlining the problematic nature of case marking in L2 speakers (e.g. Hopp, 2010; Jackson, 2007; Rankin, 2014), however these studies mainly contrasted subject and object case marking and not the case marking of two objects. It is also related to the concept of cue perceivability within the Competition Model that states that cues that are easier to perceive are acquired earlier. The distinction between the dative determiner *dem* and the accusative determiner *den* might be too hard to perceive for L2 speakers, leading to underspecification in the L2 grammar.

L2 speakers have problems with gradient acceptability, but exhibit the same acceptability-frequency gap as L1 speakers.

As stated previously, native speakers in this thesis perceived reliable differences in acceptability between canonical and non-canonical orders. In the case of Study 2 this difference was small, but in line with previous findings that had established a gradient decline in perceived acceptability for scrambled orders (Keller, 2000; Kempen & Harbusch, 2005, 2008). L2 speakers did not seem to be sensitive for small differences in acceptability; instead they evaluated both order variations as equally acceptable (see Study 2). However, L2 speakers judged non-canonical orders as reliably less acceptable than canonical orders, if there was no gradience in the L1 data (see Study 3). I suggest that L2 speakers operate with fewer distinctions of acceptability, e.g. correct – incorrect – I don't know, than native speakers who possibly operate with distinctions like best – better – good – ok – bad – worse – worst, and that these few categories make finer-grained distinctions difficult in L2 speakers. The ability of L2 speakers to accurately judge low frequency word orders that was also attested by Hopp (2005) goes against usage-based or experience-based models of language acquisition. According to these models, the low frequency of scrambled orders should increase the difference in acceptability between a familiar and less familiar order, assuming that L2 speakers are aware that one order is more frequent. Instead, L2 speakers exhibited the same acceptability-frequency gap that has been found in native speakers.

There is no true optionality in an L1, and optionality in an L2 is the result of underspecification in the L2 interlanguage grammar.

The two orders of particle and direct object in Norwegian particle verbs have been claimed by the Norwegian Reference Grammar (Faarlund et al., 1997) to be free variations of each other for short NP objects. However, the findings of Study 3 showed clear preferences in acceptability and faster processing for the order in which verb and particle were adjacent, avoiding a filler-gap dependency. The same applies to the results of Study 2: the differences in acceptability between the canonical unscrambled and the non-canonical scrambled order might be small, but they are reliable and speak against optionality in the placement of the two objects. There might be many different factors contributing to the decision which order is the suitable one in a certain context and native speakers might not be able to accurately describe the reasons for the decision for or against a certain order. Preferences in native speaker intuitions and differences in frequency of occurrence between orders suggest that phenomena like particle placement do follow rules and are not subject to free variation. The optionality that can be found in L2 speakers on the other hand is caused by an underspecification of ordering factors in the L2 interlanguage grammar. If it is not specified in the L2 grammar that, for example, DAT > ACC is preferred over ACC > DAT in sentences without context, then the Dative constraint cannot be applied, leading to optionality in the ordering of the direct and indirect objects.

Claims by descriptive grammar and those based on non-experimental judgments of acceptability need to be more thoroughly tested by experiments.

Study 1 and 3 on Norwegian were conducted to investigate claims by the Norwegian Reference Grammar regarding the acceptability of particle shift for long object NPs and the use of the order of main verb and NP2 to identify object topicalizations. None of the claims made by the NRG could be fully supported by experimental evidence. The study on scrambling in Russian by Kallestinova (2007) also did not support the wide-spread theoretical assumption that DAT > ACC is the preferred order in context-free ditransitive sentences. A discrepancy between theoretical claims and language reality is problematic for psycholinguistic research as experimental manipulations are derived from language theory. Studies conducted in languages like English, German or the Romance languages have helped to adapt claims of descriptive grammar that could not be experimentally supported to the new findings. Smaller, less-researched languages, like Norwegian, are still in a stage in which many claims by descriptive grammar have not been tested yet and might actually be incorrect. Moving away from judgments by small groups of speakers – who are often themselves linguists as already addressed in Section 2.4.2 – towards a larger scale, experimental collection of judgment data as recommended by Gibson & Fedorenko (2013) and Gibson, Piantadosi, & Fedorenko (2013), could help to identify problematic claims. In a way, the Norwegian experiments in this thesis served as pilot studies for future research verifying these (and other) claims of descriptive Norwegian grammar. Future research will have to follow up to provide more insights, for example, into L1 speakers' highly variable ability to correctly identify the agent in globally unambiguous object topicalization sentences.

10 Bibliography

Aaronson, D., & Scarborough, H. S. (1976). Performance theories for sentence coding: Some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(1), 56–70. http://doi.org/10.1037/0096-1523.2.1.56

Abrahamsson, N. (2012). Age of Onset and Nativelike L2 Ultimate Attainment of Morphosyntactic and Phonetic Intuition. *Studies in Second Language Acquisition*, *34*, 187–214.

Abrahamsson, N., & Hyltenstam, K. (2008). The Robustness of Aptitude Effects in Near-Native Second Language Acquisition. *Studies in Second Language Acquisition*, *30*(4), 481. http://doi.org/10.1017/S027226310808073X

Åfarli, T. A. (1985). Norwegian Verb Particle Constructions as Causative Constructions. *Nordic Journal of Linguistics*, *8*(1), 75–98. http://doi.org/10.1017/S0332586500001268

Åfarli, T. A., & Eide, K. M. (2003). *Norsk generativ syntaks*. Oslo: Novus forlag.

Alejo González, R. (2010). Making sense of phrasal verbs: A cognitive linguistic account of L2 learning. *AILA Review*, *23*, 50–71. http://doi.org/10.1075/aila.23.04ale

Altmann, G. T. ., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238. http://doi.org/10.1016/0010-0277(88)90020-0

Anderssen, M., Bentzen, K., Rodina, Y., & Westergaard, M. (2010). The Acquisition of Apparent Optionality: Word Order in Subject and Object Shift Constructions in Norwegian. In M. Anderssen, K. Bentzen, & M. Westergaard (Eds.), *Variation in the Input* (Vol. 39, pp. 241–270). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-90-481-9207-6

Andréasson, M. (2009). Pronominal object shift – not just a matter of shifting or not. *Working Papers in Scandinavian Syntax*, *84*(August), 1–20.

Askedal, J. O. (1982). Om preposisjonens syntaktiske status i de sakalte « transitive verbalgrupper » i norsk. *Nordic Journal of Linguistics*, *5*(1), 133–162.

Askedal, J. O. (1984). On Extraposition in German and Norwegian. Towards a Contrastive Analysis. *Nordic Journal of Linguistics*, *7*(2), 83–113. http://doi.org/10.1017/S0332586500001165

Bader, M. (2000). On Reanalysis: Evidence from German. In B. Hemforth & L. Konieczny (Eds.), *German Sentence Processing* (pp. 187–246). Kluwer Academic Publishers.

Bader, M., Bayer, J., Hopf, J. M., & Meng, M. (1996). Case-assignment in processing German verb-final clauses. In *Proceedings of the NELS 26 Sentence Processing Workshop* (pp. 1–25). MIT Occasional Papers in Linguistics.

Bader, M., & Häussler, J. (2010). Word order in German : A corpus study. *Lingua*, *120*, 717–762. http://doi.org/10.1016/j.lingua.2009.05.007

Bader, M., & Meng, M. (1999). Subject-Object Ambiguities in German Embedded Clauses: An Across-the-Board Comparison. *Journal of Psycholinguistic Research*, *28*(2), 121–143.

Bailyn, J. F. (2010). What's Inside VP? New (and old) Evidence from Russian. In *Formal Approaches to Slavic Linguistics* (Vol. 18, pp. 21–37).

Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude Estimation of Linguistic Acceptibility. *Language*, *72*(1), 32–68.

Baten, K., & De Cuypere, L. (2014). The dative alternation in L2 German? Conceptualization transfer from L1 Dutch. *Vigo International Journal of Applied Linguistics*, *11*(1), 9–40.

Bentzen, K. (2007a). *Order and Structure in Embedded Clauses in Northern Norwegian*.

Bentzen, K. (2007b). The degree of verb movement in embedded clauses in three varieties of Norwegian. *Nordlyd*, *34*, 127–146.

Berwick, R. C., Pietroski, P., Yankama, B., & Chomsky, N. (2011). Poverty of the stimulus revisited. *Cognitive Science*, *35*(7), 1207–42. http://doi.org/10.1111/j.1551-6709.2011.01189.x

Biskup, P. (2006). Scrambling in Czech: Syntax, semantics, and information structure. In *Proceedings of Northwest Linguistics Conference* (Vol. 21, pp. 1–15). Retrieved from http://www.uni-leipzig.de/~biskup/Scrambling in Czech syntax, semantics, and information structure.pdf

Blais, M.-J., & Gonnerman, L. M. (2013). Explicit and implicit semantic processing of verb – particle constructions by French – English bilinguals. *Bilingualism: Language and Cognition*, *16*(4), 829–846. http://doi.org/10.1017/S1366728912000673

Bley-Vroman, R. (1990). The Logical Problem of Foreign Language Learning. *Linguistic Analysis*, *20*(1–2), 3–49.

Bohnacker, U. (2006a). Placing verbs and particles in non-native German and Swedish. *Working Papers in Scandinavian Syntax*, *78*, 77–127.

Bohnacker, U. (2006b). When Swedes begin to learn German: from V2 to V2. *Second Language Research*, *22*(4), 443–486. http://doi.org/10.1191/0267658306sr2750a

Bohnacker, U. (2007). On the "Vulnerability" of Syntactic Domains in Swedish and German. *Language Acquisition*, *14*(1), 31–73. http://doi.org/10.2307/20462479

Bohnacker, U. (2010). The clause-initial position in L2 Swedish declaratives: Word order variation and discourse pragmatics. *Nordic Journal of Linguistics*, *33*(2), 105–143. http://doi.org/10.1017/S033258651000017X

Bohnacker, U., & Rosén, C. (2008). The clause-initial position in L2 German declaratives. Transfer of Information. *Studies in Second Language Acquisition*, *30*(4), 511. http://doi.org/10.1017/S0272263108080741

Bornkessel, I. D., & Schlesewsky, M. (2006a). The Extended Argument Dependency Model : A Neurocognitive Approach to Sentence Comprehension Across Languages, *113*(4), 787–821. http://doi.org/10.1037/0033-295X.113.4.787

Bornkessel, I. D., & Schlesewsky, M. (2006b). The Role of Contrast in the Local Licensing of Scrambling in German: Evidence from Online Comprehension. *Journal of Germanic Linguistics*, *18*(1), 1–43. http://doi.org/10.1017/S1470542706000018

Bornkessel, I. D., Schlesewsky, M., & Friederici, A. D. (2002). Grammar overrides frequency: evidence from the online processing of flexible word order. *Cognition*, *85*(2), B21–B30. http://doi.org/10.1016/S0010-0277(02)00076-8

Bornkessel, I. D., Schlesewsky, M., & Friederici, A. D. (2003). Eliciting thematic reanalysis effects: The role of syntax-independent information during parsing. *Language and Cognitive Processes*, *18*(3), 269–298.

http://doi.org/10.1080/01690960244000018
Bornkessel, I. D., 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), 221–33. http://doi.org/10.1016/j.neuroimage.2005.01.032
Bošković, Ž., & Takahashi, D. (1998). Scrambling and Last Resort. *Linguistic Inquiry*, *29*(3), 347–366.
Boston, M. F., Hale, J. T., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, *26*(3), 301–349. http://doi.org/10.1080/01690965.2010.492228
Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, *B*(26), 211–234.
Bresnan, J. (2001). *Lexical-Functional Syntax*. Blackwell.
Broekhuis, H. (2008). *Derivations and Evaluations: Object Shift in the Germanic Languages*. Berlin: Mouton de Gruyter.
Cappelle, B., Shtyrov, Y., & Pulvermüller, F. (2010). Heating up or cooling up the brain? MEG evidence that phrasal verbs are lexical units. *Brain and Language*, *115*(3), 189–201. http://doi.org/10.1016/j.bandl.2010.09.004
Chomsky, N. (1957). *Syntactic Structures*. ’s-Gravenhage: Mouton & Co.
Chomsky, N. (1981). *Lectures on government and binding*. Berlin: Mouton de Gruyter.
Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407. http://doi.org/10.1006/cogp.2001.0752
Clahsen, H., & Felser, C. (2006a). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*, 107–126.
Clahsen, H., & Felser, C. (2006b). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42. http://doi.org/10.1017/S0142716406060024
Clahsen, H., & Felser, C. (2006c). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–70. http://doi.org/10.1016/j.tics.2006.10.002
Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological Structure in Native and Nonnative Language Processing. *Language Learning*, *60*(1), 21–43.
Clifton, C., & Frazier, L. (2004). Should given information come before new? Yes and no. *Memory & Cognition*, *32*(6), 886–895. http://doi.org/10.3758/BF03196867
Collins, C., & Thráinsson, H. (1996). VP-Internal Structure and Object Shift in Icelandic. *Linguistic Inquiry*, *27*(3), 391–444.
Corver, N., & van Riemsdijk, H. (1994). Introduction: approaches to and properties of scrambling. In *Studies on Scrambling: Movement and non movement approaches to free word order phenomena* (pp. 1–16). Berlin: Mouton de Gruyter.
Coughlin, C. E., & Tremblay, A. (2012). Proficiency and working memory based explanations for nonnative speakers’ sensitivity to agreement in sentence processing. *Applied Psycholinguistics*, *34*(3), 615–646. http://doi.org/10.1017/S0142716411000890
Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs - a case for contrastive analysis. *Studies in Second Language Acquisition*, *7*(1), 73–79.

http://doi.org/10.1093/applin/19.1.24
De Cuypere, L., De Coster, E., & Baten, K. (2014). The acquisition of the English dative alternation by Russian Foreign Language Learners. In *Phrasis (Gent): Studies in Language and Literature* (Vol. 2, pp. 187–212).
Dehé, N. (2005). The optimal placement of up and ab - A comparison. *Journal of Comparative Germanic Linguistics*, *8*(3), 185–224. http://doi.org/10.1007/s10828-004-1686-9
Dehé, N., Jackendoff, R., McIntyre, A., & Urban, S. (2002). Introduction. In N. Dehé, R. Jackendoff, A. McIntyre, & S. Urban (Eds.), *Verb-Particle Explorations* (pp. 1–20). Berlin: Mouton de Gruyter.
den Dikken, M. (1995). *Particles: On the Syntax of Verb-particle, Triadic, and Causative Constructions*. New York, NY: Oxford University Press.
Desmet, T., De Baecke, C., Drieghe, D., Brysbaert, M., & Vonk, W. (2006). Relative clause attachment in Dutch: On-line comprehension corresponds to corpus frequencies when lexical variables are taken into account, (March 2015), 453–485. http://doi.org/10.1080/01690960400023485
Domínguez, L., & Arche, M. J. (2008). Optionality in L2 grammars: the acquisition of SV/VS contrast in Spanish. In H. Chan, H. Jacob, & E. Kapia (Eds.), *Proceedings of the 32nd Annual Boston University Conference on Language Development* (pp. 96–107). Somerville, MA, USA: Cascadilla. Retrieved from http://gala.gre.ac.uk/1447/
Dvořák, V. (2010). On the Syntax of Ditransitive Verbs in Czech. In W. Browne (Ed.), *Formal Approaches to Slavic Linguistics 18: The Second Cornell Meeting 2009* (pp. 161–177). Ann Arbor, MI: Michigan Slavic Publications. Retrieved from http://www.rci.rutgers.edu/~dvorakov/docs/FASL18-article.pdf
Erdocia, K., Laka, I., Mestres-Missé, A., & Rodríguez-Fornells, A. (2009). Syntactic complexity and ambiguity resolution in a free word order language: behavioral and electrophysiological evidences from Basque. *Brain and Language*, *109*(1), 1–17. http://doi.org/10.1016/j.bandl.2008.12.003
Faarlund, J. T., Lie, S., & Vannebo, K. I. (1997). *Norsk referansegrammatikk*. (J. T. Faarlund, S. Lie, & K. I. Vannebo, Eds.). Columbia University Press.
Fanselow, G. (1990). Scrambling as NP-Movement. In G. Grewendorf & W. Sternefeld (Eds.), *Scrambling and Barriers* (pp. 113–140). Amsterdam: Benjamins.
Fanselow, G. (2001). Features , theta-Roles, and Free Constituent Order. *Linguistic Inquiry*, *32*(3), 405–437.
Fanselow, G., & Frisch, S. (2006). Effects of Processing Difficulty on Judgments of Acceptability. In G. Fanselow, C. Féry, R. Vogel, & M. Schlesewsky (Eds.), *Gradience in Grammar. Generative perspectives* (Oxford Lin, pp. 291–316). Oxford, New York: Oxford University Press.
Fedorenko, E., Woodbury, R., & Gibson, E. (2013). Direct evidence of memory retrieval as a source of difficulty in non-local dependencies in language. *Cognitive Science*, *37*(2), 378–94. http://doi.org/10.1111/cogs.12021
Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. http://doi.org/10.1016/S0010-0285(03)00005-7
Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. http://doi.org/10.1111/1467-8721.00158
Ferreira, F., Engelhardt, P. E., & Jones, M. W. (2009). Good Enough Language

Processing: A Satisficing Approach. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, (1), 413–418.
Ferreira, F., & Patson, N. D. (2007). The “Good Enough” Approach to Language Comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. http://doi.org/10.1111/j.1749-818X.2007.00007.x
Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: insights from ERPs and fMRI. *Journal of Psycholinguistic Research*, *30*(3), 321–338. http://doi.org/10.1023/A:1010447102554
Fodor, J. D. (1989). Empty Categories in Sentence Processing. *Language and Cognitive Processes*, *4*(3–4), 155–209.
Forster, K. I., & Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *35*(1), 116–124. http://doi.org/10.3758/BF03195503
Foucart, A., & Frenck-Mestre, C. (2010). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*(3), 379–399. http://doi.org/10.1017/S136672891000012X
Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*(1), 226–248. http://doi.org/10.1016/j.jml.2011.07.007
Frazier, L. (1987). Syntactic Processing : Evidence from Dutch. *Natural Language and Linguistic Theory*, *5*(4), 519–559.
Frazier, L., & Clifton, C. (1997). Construal : Overview , Motivation , and Some New Evidence, *26*.
Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325. http://doi.org/10.1016/0010-0277(78)90002-1
Fretheim, T. (1992). Themehood, Rhemehood and Norwegian Focus Structure. *Folia Linguistica*, *26*(C), 111–150. http://doi.org/10.1515/flin.1992.26.1-2.111
Friederici, A. D., & Frisch, S. (2000). Verb Argument Structure Processing: The Role of Verb-Specific and Argument-Specific Information. *Journal of Memory and Language*, *43*(3), 476–507. http://doi.org/10.1006/jmla.2000.2709
Georgala, E. (2011a). Scrambling Verb-Final Languages and the Underlying Order of Objects in Ditransitive Constructions. *University of Pennsylvania Working Papers in Linguistics*, *17*(1), 105–114.
Georgala, E. (2011b). Why German Is Not an Exception to the Universal < IO , DO > Base Order of Double Object Constructions. In M. Byram Washburn, K. McKinney-Bock, E. Varis, A. Sawyer, & B. Tomaszewicz (Eds.), *Proceedings of the 28th West Coast Conference on Formal Linguistics* (pp. 96–105). Sommerville, Mass: Cascadilla Proceedings Project.
Gerth, S., Otto, C., Felser, C., & Nam, Y. (2015). Strength of garden-path effects in native and non-native speakers’ processing of object-subject ambiguities. *International Journal of Bilingualism*, 1–20. http://doi.org/10.1177/1367006915604401
Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. http://doi.org/10.1016/S0010-0277(98)00034-1
Gibson, E. (2000). The Dependency Locality Theory : A Distance -Based Theory

of Linguistic Complexity. In Y. Miyashita, A. Marantz, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Gibson, E., & Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, *1*(2), 88–124.

Gibson, E., Piantadosi, S. T., & Fedorenko, E. (2013). Quantitative methods in syntax / semantics research: A response to Sprouse and Almeida (2013). *Language and Cognitive Processes*, *28*(2), 229–240.

Gibson, E., Piantadosi, S. T., & Fedorenko, K. (2011). Using Mechanical Turk to Obtain and Analyze English Acceptability Judgments. *Language and Linguistics Compass*, *5*(8), 509–524. http://doi.org/10.1111/j.1749-818X.2011.00295.x

Godfroid, A., Loewen, S., Jung, S., Park, J.-H., Gass, S., & Ellis, R. (2015). Timed And Untimed Grammaticality Judgments Measure Distinct Types Of Knowledge. Evidence from Eye-Movement Patterns. *Studies in Second Language Acquisition*, *37*(c), 269–297. http://doi.org/10.1017/S0272263114000850

Gracanin-Yuksek, M. (2006). Double object construction in Croatian: Arguments against Appl0. In R. Compton, M. Goledzinowska, & U. Savchenko (Eds.), *Formal approaches to Slavic linguistics: The Toronto meeting* (pp. 94–112). Ann Arbor, MI: Michigan Slavic Publications. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Double+Object+Construction+in+Croatian:+Arguments+Against+Appl#0

Gries, S. T. (2002). The influence of processing on syntactic variation: Particle placement in English. In N. Dehé, R. Jackendoff, A. McIntyre, & S. Urban (Eds.), *Verb-Particle Explorations* (pp. 269–288). Berlin: Mouton de Gruyter.

Haddican, B., & Johnson, D. E. (2012). Effects on the particle verb alternation across English dialects. *Selected Papers from NWAV 40*, *18*(2), 31-.

Haider, H. (2006). Mittelfeld phenomena (scrambling in Germanic). In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax* (pp. 204–274).

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–79. http://doi.org/10.1126/science.298.5598.1569

Havik, E., Roberts, L., van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing Subject-Object Ambiguities in the L2 : A Self-Paced Reading Study With German L2 Learners of Dutch. *Language Learning*, *59*(1), 73–112.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, J. A. (1999). Processing Complexity and Filler-Gap Dependencies across Grammars. *Language*, *75*(2), 244–285.

Hawkins, J. A. (2011). Discontinuous Dependencies in Corpus Selections: Particle Verbs and their Relevance for Current Issues in Language Processing. In E. M. Bender & J. E. Arnold (Eds.), *Language from a cognitive perspective: Grammar, usage, and processing* (pp. 243–265). Stanford: CSLI Publications.

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB – eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, *62*(1), 10–20. http://doi.org/10.1026/0033-3042/a000029

Heylen, K. (2005). A Quantitative Corpus Study of German Word Order Variation. In S. Kepser & M. Reis (Eds.), *Linguistic Evidence-Empirical, Theoretical, and Computational Perspectives* (pp. 241–263). Berlin: Mouton de Gruyter.

Heylen, K., & Speelman, D. (2003). A corpus-based analysis of word order variation. The order of verb arguments in the German middle field. In *Proceedings of the Corpus Linguistics 2003 conference* (pp. 320–329). Lancaster: Lancaster University Papers.

Holmberg, A. (1986). *Word Order and syntactic features in the Scandinavian languages and English*. University of Stockholm.

Holmberg, A., & Platzack, C. (2005). The Scandinavian Languages. In G. Cinque & R. S. Kayne (Eds.), *The Oxford Handbook of Comparative Syntax* (pp. 420–458). Oxford: Oxford University Press.

Hopp, H. (2005). Constraining second language word order optionality: scrambling in advanced English–German and Japanese–German interlanguage. *Second Language Research*, *21*(1), 34–71. http://doi.org/10.1191/0267658305sr246oa

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, *22*(3), 369–397. http://doi.org/10.1191/0267658306sr272oa

Hopp, H. (2007). *Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing*.

Hopp, H. (2009). The syntax-discourse interface in near-native L2 acquisition : Off-line and on-line performance. *Bilingualism : Language and Cognition*, *12*(4), 463–483. http://doi.org/10.1017/S1366728909990253

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901–931. http://doi.org/10.1016/j.lingua.2009.06.004

Hopp, H. (2015). Individual differences in the second language processing of object–subject ambiguities. *Applied Psycholinguistics*, *36*(2), 129–173. Retrieved from http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=8888092

Hosono, M. (2010a). Scandinavian Object Shift as the Cause of Downstep. *Working Papers in Scandinavian Syntax*, *85*, 1–36.

Hosono, M. (2010b). Why Object Shift does not exist in Övdalian. *Working Papers in Scandinavian Syntax*, *86*, 57–80.

Hróarsdóttir, T. (2008). Verb particles in OV/VO word order in Older Icelandic. *Working Papers in Scandinavian Syntax*, *82*(2008), 43–81.

Institute, G. (2010). “Testen Sie Ihr Deutsch/ Test your German.” Retrieved from http://www.goethe.de/cgi-bin/einstufungstest/einstufungstest.pl

Jackson, C. N. (2007). The use and non-use of semantic information, word order, and case markings during comprehension by L2 learners of German. *Modern Language Journal*, *91*, 418–432. http://doi.org/10.1111/j.1540-4781.2007.00588.x

Jackson, C. N. (2008). Proficiency Level and the Interaction of Lexical and Morphosyntactic Information During L2 Sentence Processing. *Language Learning*, *58*(4), 875–909.

Jackson, C. N. (2010). The processing of subject-object ambiguities by English Dutch L2 learners of German. In B. Van Patten & J. Jegerski (Eds.), *Research in Second Language Processing and Parsing* (pp. 207–230).

Jackson, C. N., & Roberts, L. (2010). Animacy affects the processing of subject–object ambiguities in the second language: Evidence from self-paced reading with German second language learners of Dutch. *Applied Psycholinguistics*, *31*(4), 671–691. http://doi.org/10.1017/S0142716410000196

Jacob, G., & Felser, C. (2015). Reanalysis and semantic persistence in native and non-native garden-path recovery. *The Quarterly Journal of Experimental Psychology*, 1–19. http://doi.org/10.1080/17470218.2014.984231

Jäschke, K., & Plag, I. (2016). The Dative Alternation in German-English Interlanguage. *Studies in Second Language Acquisition*, *38*, 1–66.

Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 20–49). New York, NY: Routledge.

Jin, F., Åfarli, T. A., & van Dommelen, W. A. (2007). Endstate L2 Competence: Does Perception Parallel Production? In A. Belikova, L. Meroni, & M. Umeda (Eds.), *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America* (pp. 200–211). Somerville: Cascadilla Proceedings Project.

Jörgensen, N. (1976). *Meningsbyggnaden i talad svenska*. Lund: Studentlitteratur.

Josefsson, G. (2010). *Object shift and optionality. An intricate interplay between syntax, prosody and information structure*. *Working Papers in Scandinavian Syntax* (Vol. 86).

Juffs, A. (2001). Psycholinguistically oriented second language research. *Annual Review of Applied Linguistics*, *21*, 207–221. http://doi.org/10.1017/S0267190501000125

Juffs, A., & Harrington, M. (1996). Garden Path Sentences and Error Data in Second Language Sentence Processing. *Language Learning*, *46*(2), 283–326.

Junghanns, U., & Lenertová, D. (2007). *Realizing argument structure of ditransitives in Slavic*.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, *87*(January), 329–354. http://doi.org/10.1037/0033-295X.87.4.329

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology. General*, *111*(2), 228–238. http://doi.org/10.1037/0096-3445.111.2.228

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. http://doi.org/10.1016/j.cognition.2004.01.002

Kallestinova, E. D. (2007). *Aspects of word order in Russian*.

Keller, F. (2000). *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality*. Retrieved from http://hdl.handle.net/1842/744

Kempe, V., & MacWhinney, B. (1998). The Acquisition of Case Marking By Adult Learners of Russian and German. *Studies in Second Language Acquisition*, *20*(4), 543–587. http://doi.org/10.1017/S0272263198004045

Kempe, V., & MacWhinney, B. (1999). Processing of Morphological and Semantic Cues in Russian and German. *Language and Cognitive Processes*, *14*(2), 129–171.

Kempen, G., & Harbusch, K. (2003). An artificial opposition between grammaticality and frequency: Comment on Bornkessel, Schlesewsky, and

Friederici (2002). *Cognition*, *90*(2), 205–210. http://doi.org/10.1016/S0010-0277(03)00145-8
Kempen, G., & Harbusch, K. (2004a). A corpus study into word order variation in German subordinate clauses : Animacy affects linearization independently of grammatical function assignment. In T. Pechmann & C. Habel (Eds.), *Multidisciplinary approaches to language production* (pp. 173–182). Berlin: Mouton de Gruyter.
Kempen, G., & Harbusch, K. (2004b). Generating natural word orders in a semi–free word order language: Treebank-based linearization preferences for German. *Computational Linguistics and Intelligent Text Processing*, *2945*, 350–354.
Kempen, G., & Harbusch, K. (2004c). How flexible is constituent order in the midfield of German subordinate clauses? A corpus study revealing unexpected rigidity. In *Proceedings of the International Conference on Linguistic Evidence* (pp. 29–31). Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=3626049423831512O7related:Z3iMIEY7CAUJ
Kempen, G., & Harbusch, K. (2005). The relationship between grammaticality ratings and corpus frequencies: A case study into word order variability in the midfield of German clauses. In S. Kepser & M. Reis (Eds.), *Linguistic Evidence-Empirical, Theoretical, and Computational Perspectives* (pp. 1–22). Berlin: Mouton de Gruyter. Retrieved from http://books.google.com/books?hl=en&lr=&id=tKbrYoZ7zCEC&oi=fnd&pg=PA329&dq=The+relationship+between+grammaticality+ratings+and+corpus+frequencies+:+A+case+study+into+word+order+variability+in+the+midfield+of+German+clauses&ots=cqbk83Xz8G&sig=8Cnwb2JFYRw54
Kempen, G., & Harbusch, K. (2008). Comparing Linguistic Judgments and Corpus Frequencies as Windows on Grammatical Competence: A Study of Argument Linearization in German Clauses. *Language, Context, & Cognition: The Discourse Potential of Underspecified Structures*, 179–192.
Kilborn, K., & Ito, T. (1989). Sentence Processing Strategies in Adult Bilinguals. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence comprehension* (pp. 257–291). Cambridge: Cambridge University Press.
King, J. W., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602. http://doi.org/10.1016/0749-596X(91)90027-H
King, T. H. (1995). *Configuring topic and focus in Russian*. Stanford University.
Kiss, K. É. (1994). Scrambling as the base-generation of random complement order. In N. Corver & H. van Riemsdijk (Eds.), *Studies on Scrambling: Movement and non-movement approaches to free word-order phenomena* (pp. 221–257). Berlin: Mouton de Gruyter.
Konieczny, L. (2000). Locality and Parsing Complexity, *29*(6), 627–645.
Konieczny, L., Hemforth, B., Scheepers, C., & Strube, G. (1997). The Role of Lexical Heads in Parsing: Evidence from German. *Language and Cognitive Processes*, *12*(2–3), 307–348. http://doi.org/10.1080/016909697386871
Köpcke, K.-M., & Zubin, D. A. (1979). Gibt es außerhalb der logisch-grammatischen Struktur eine perzeptuelle Satzgliederung? In *Sprachstruktur, Individuum und Gesellschaft: Akten des 13. Linguistischen Kolloquiums* (pp. 255–269).
Kosta, P. (2006). On Free Word Order Phenomena in Czech as Compared to

German: Is Clause Internal Scrambling A-Movement, A-Bar-Movement or Is It Base Generated? *Zeitschrift Für Slawistik*, *51*(3), 306–320. http://doi.org/10.1524/slaw.2006.51.3.306

Kristensen, L. B., Engberg-Pedersen, E., Højlund Nielsen, A., & Wallentin, M. (2013). The influence of context on word order processing - An fMRI study. *Journal of Neurolinguistics*, *26*(1), 73–88. http://doi.org/10.1016/j.jneuroling.2012.05.001

Kristensen, L. B., Engberg-Pedersen, E., & Poulsen, M. (2014). Context Improves Comprehension of Fronted Objects. *Journal of Psycholinguistic Research*, *43*(2), 125–40. http://doi.org/10.1007/s10936-013-9241-y

Kurz, D. (2000). A Statistical Account on Word Order Variation in German. In A. Abeillé, T. Brants, & H. Uszkoreit (Eds.), *Proceedings of the COLING Workshop on Linguistically Interpreted Corpora* (pp. 35–41). Luxembourg.

Larson, R. K. (1988). On the Double Object Construction. *Linguistic Inquiry*, *19*(3), 335–391. http://doi.org/10.2307/25164901

Laufer, B., & Eliasson, S. (2008). What Causes Avoidance in L2 Learning. *Studies in Second Language Acquisition*, *15*(1), 35. http://doi.org/10.1017/S0272263100011657

Lewis, R. L., & Vasishth, S. (2005). An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval, *29*, 375–419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–54. http://doi.org/10.1016/j.tics.2006.08.007

Liao, Y., & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of chinese learners of english. *Language Learning*, *54*(June), 193–226. http://doi.org/10.1111/j.1467-9922.2004.00254.x

Liberman, M. (2005). Syntax is a disturbance in the there. Retrieved from http://itre.cis.upenn.edu/~myl/languagelog/archives/002178.html

Lobanova, A. (2011). The role of prominence scales for the disambiguation of grammatical functions in Russian. *Russian Linguistics*, *35*, 125–142. http://doi.org/10.1007/s11185-010-9066-3

Lohse, B., Hawkins, J. A., & Wasow, T. (2004). Domain minimization in English verb-particle constructions. *Language*, *80*(2), 238–261.

MacDonald, M. C. (1997). Lexical Representations and Sentence Processing : An Introduction. *Language and Cognitive Processes*, *12*(2–3), 121–136.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*(1), 56–98. http://doi.org/10.1016/0010-0285(92)90003-K

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical Nature of Syntactic Ambiguity Resolution, *101*(4), 676–703.

MacWhinney, B. (1997). Second language acquisition and the Competition Model. *Tutorials in Bilingualism: Psycholinguistic Perspectives*, 113–142.

MacWhinney, B. (2005). Extending the Competition Model. *International Journal of Bilingualism*, *9*(1), 69–84. http://doi.org/10.1177/13670069050090010501

MacWhinney, B., & Bates, E. A. (1989). *The crosslinguistic study of sentence processing*. New York, NY: Cambridge University Press.

MacWhinney, B., Bates, E. A., & Kliegl, R. (1984). Cue Validity and Sentence Interpretation in English , German , and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*, 127–150.

Mak, W. M., Vonk, W., & Schriefers, H. J. (2008). Discourse structure and

relative clause processing. *Memory & Cognition*, *36*(1), 170–181. http://doi.org/10.3758/MC.36.1.170
Marefat, H. (2005). The impact of information structure as a discourse factor on the acquisition of dative alternation by L2 learners. *Studia Linguistica*, *59*(1), 66–82.
Marvin, T., & Stegovec, A. (2012). On the syntax of ditransitive sentences in Slovenian. *Acta Linguistica Hungarica*, *59*(1–2), 177–203. Retrieved from http://www.akademiai.com/index/T724251Q4Q4783MW.pdf
Matlock, T., & Heredia, R. R. (2002). Understanding phrasal verbs in monolinguals and bilinguals. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual Sentence Processing* (pp. 251–274). Elsevier Science B.V. http://doi.org/10.1016/S0166-4115(02)80014-0
McClelland, J. L., John, M. St., & Taraban, R. (1989). Sentence Comprehension : A Parallel Distributed Processing Approach. *Language and Cognitive Processes*, *4*(3–4), S1287–S1335.
McDonald, J. L. (1987). Assigning linguistic roles: The influence of conflicting cues. *Journal of Memory and Language*, *26 1*, 100–117. http://doi.org/10.1016/0749-596X(87)90065-9
McDonald, J. L. (1989). The Acquisition of Cue-Category Mappings. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence comprehension* (pp. 375–396). Cambridge: Cambridge University Press.
Meng, M., & Bader, M. (2000a). *Mode of Disambiguation and Garden-Path Strength: An Investigation of Subject-Object Ambiguities in German. Language and Speech* (Vol. 43). http://doi.org/10.1177/00238309000430010201
Meng, M., & Bader, M. (2000b). Ungrammaticality detection and garden path strength : Evidence for serial parsing. *Language and Cognitive Processes*, *15*(6), 615–666.
Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, *24*(6), 469–488.
Mitchell, D. C., & Green, D. W. (1978). The effects of context and content on immediate processing in reading, (September 2015). http://doi.org/10.1080/14640747808400689
Müller, G. (1999). Optimality, markedness, and word order in German. *Linguistics*, *37*(1999), 777–818. http://doi.org/10.1515/ling.37.5.777
Müller, S. (2002). *Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German. Habilitationsschrift, Universität des Saarlandes, Saarbrücken*. http://doi.org/10.1353/lan.0.0067
Mykhaylyk, R., Rodina, Y., & Anderssen, M. (2013). Ditransitive constructions in Russian and Ukrainian: Effect of givenness on word order. *Lingua*, *137*, 271–289. http://doi.org/10.1016/j.lingua.2013.10.001
Neeleman, A. (2002). Particle placement. In N. Dehé, R. Jackendoff, A. McIntyre, & S. Urban (Eds.), *Verb-Particle Explorations* (pp. 141–164). Berlin: Mouton de Gruyter.
Neeleman, A., & Titov, E. (2009). Focus, Contrast and Stress in Russian. *Linguistic Inquiry*, *40*(3), 514–524.
Øvrelid, L. (2004). Disambiguation of syntactic functions in Norwegian: modeling variation in word order interpretations conditioned by animacy

and definiteness. In *Proceedings of the 20th Scandinavian Conference of Linguistics* (pp. 1–17).

Pan, H.-Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*, *121*(2), 221–236. http://doi.org/10.1016/j.lingua.2010.08.003

Papadopoulou, D. (2005). Reading-time studies of second language ambiguity resolution. *Second Language Research*, *21*(2), 98–120. http://doi.org/10.1191/0267658305sr254oa

Papadopoulou, D., Varlokosta, S., Spyropoulos, V., Kaili, H., Prokou, S., & Revithiadou, A. (2011). Case morphology and word order in second language Turkish: Evidence from Greek learners. *Second Language Research*, *27*(2), 173–204. http://doi.org/10.1177/0267658310376348

Pappert, S., Schließer, J., Janssen, D. P., & Pechmann, T. (2007). Corpus and psycholinguistic investigations of linguistic constraints on German object order. In A. Späth (Ed.), *Interfaces and interface conditions* (pp. 299–328). Berlin: Mouton de Gruyter. Retrieved from http://kar.kent.ac.uk/4409/

Park, K.-S. (2011). Information structure and dative word order in adult L2 learners. In J. Herschensohn & D. Tanner (Eds.), *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011)* (pp. 101–109). Somerville, MA: Cascadilla Proceedings Project.

Patterson, C., Trompelt, H., & Felser, C. (2014). The online application of binding condition B in native and non-native pronoun resolution. *Frontiers in Psychology*, *5*(FEB), 1–16. http://doi.org/10.3389/fpsyg.2014.00147

Paulmann, S., Ghareeb-Ali, Z., & Felser, C. (2015). Neurophysiological Markers of Phrasal Verb Processing : Evidence From L1 and L2 Speakers. In R. R. Heredia & A. B. Cieslicka (Eds.), *Bilingual Figurative Language Processing*. Cambridge University Press.

Piai, V. M., Meyer, L., Schreuder, R., & Bastiaansen, M. C. M. (2013). Sit down and read on: Working memory and long-term memory in particle-verb processing. *Brain and Language*, *127*(2), 296–306. http://doi.org/10.1016/j.bandl.2013.09.015

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 940–961. http://doi.org/10.1037/0278-7393.24.4.940

Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2000). Ambiguity Resolution in Sentence Processing: Evidence against Frequency-Based Accounts. *Journal of Memory and Language*, *43*(3), 447–475. http://doi.org/10.1006/jmla.2000.2708

Pienemann, M., & Håkansson, G. (2007). Full transfer vs. developmentally moderated transfer: a reply to Bohnacker. *Second Language Research*, *23*(4), 485–493. http://doi.org/10.1177/0267658307080332

Pollard, J., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Poulsen, M. (2012). The usefulness of the grammaticality–acceptability distinction in functional approaches to language. *Acta Linguistica Hafniensia*, *44*(1), 4–21. http://doi.org/10.1080/03740463.2011.735472

Primus, B. (1994). Grammatik und Performanz: Faktoren der Wortstellungsvariation im Mittelfeld. *Sprache Und Pragmatik*, *32*, 39–86.

Primus, B. (1998). The relative order of recipient and patient in the languages of

Europe. In A. Siewierska (Ed.), *Constituent Order in the Languages of Europe* (pp. 421–475). Berlin: Mouton de Gruyter.
Pullum, G. K. (2005). Yoda's syntax the Tribune analyzes; supply more details I will! Retrieved from http://itre.cis.upenn.edu/~myl/languagelog/archives/002173.html
Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
Ramchand, G., & Svenonius, P. (2002). The Lexical Syntax and Lexical Semantics of the Verb-Particle Construction. In L. Mikkelsen & C. Potts (Eds.), *WCCFL 21 Proceedings* (pp. 387–400). Somerville, MA: Cascadilla Press.
Rankin, T. (2014). Word order and case in the comprehension of L2 German by L1 English speakers. *EUROSLA Yearbook*, *14*(2006), 201–224. http://doi.org/10.1075/eurosla.14.08ran
Roberts, L. (2012). Psycholinguistic techniques and resources in second language acquisition research. *Second Language Research*, *28*(1), 113–127. http://doi.org/10.1177/0267658311418416
Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, *32*(2), 299–331. http://doi.org/10.1017/S0142716410000421
Rodina, Y., & Westergaard, M. (2015). Grammatical Gender in Norwegian: Language Acquisition and Language Change. *Journal of Germanic Linguistics*, *27*(2), 145–187. http://doi.org/10.1017/S1470542714000245
Roll, M., Horne, M., & Lindgren, M. (2007). Object Shift and Event-Related Brain Potentials. *Journal of Neurolinguistics*, *20*(6), 462–481. http://doi.org/10.1016/j.jneuroling.2007.04.001
Rösler, F., Pechmann, T., Streb, J., Ro, B., Hennighausen, E., & Röder, B. (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*, *176*(38), 150–176. http://doi.org/10.1006/jmla.1997.2551
Ross, J. R. (1986). *Infinite syntax!* Norwood, NJ: ABLEX.
Rydland, V., Grøver, V., & Lawrence, J. (2013). The second-language vocabulary trajectories of Turkish immigrant children in Norway from ages five to ten: the role of preschool talk exposure, maternal education, and co-ethnic concentration in the neighborhood. *Journal of Child Language*, 1–30. http://doi.org/10.1017/S0305000912000712
Sagarra, N., & Herschensohn, J. (2011). Proficiency and Animacy Effects on L2 Gender Agreement Processes During Comprehension. *Language Learning*, *61*(1), 80–116. http://doi.org/10.1111/j.1467-9922.2010.00588.x
Schlesewsky, M., & Bornkessel, I. D. (2004). On incremental interpretation: Degrees of meaning accessed during sentence comprehension. *Lingua*, *114*(9–10), 1213–1234. http://doi.org/10.1016/j.lingua.2003.07.006
Schlesewsky, M., Bornkessel, I. D., & Frisch, S. (2003). The neurophysiological basis of word order variations in German. *Brain and Language*, *86*(1), 116–128. http://doi.org/10.1016/S0093-934X(02)00540-0
Schlesewsky, M., Fanselow, G., Kliegl, R., & Krems, J. (2000). The subject preference in the processing of locally ambiguous WH-Questions in German. In *German Sentence Processing* (pp. 65–93).
Schriefers, H. J., Friederici, A. D., & Kühn, K. (1995). The Processing of Locally Ambiguous Relative Clauses in German. *Journ of Memory and Language*, *34*, 499–520.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, *12*, 40–73. http://doi.org/10.1177/026765839601200103

Siewierska, A. (1993). Syntactic weight vs. pragmatic factors and word order variation in Polish. *Journal of Linguistics*, *29*(2), 233–265. http://doi.org/10.1017/S0022226700000323

Siewierska, A., & Uhlírová, L. (1998). An overview of word order in Slavic languages. In A. Siewierska (Ed.), *Constituent Order in the Languages of Europe* (pp. 105–150). Berlin: Mouton de Gruyter.

Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *IRAL - International Review of Applied Linguistics in Language Teaching*, *45*(2), 119–139. http://doi.org/10.1515/IRAL.2007.005

Slabakova, R. (2009). L2 Fundamentals. *Studies in Second Language Acquisition*, *31*, 155–173. http://doi.org/10.1017/S0272263109090263

Slavkov, N. (2008). Formal Consequences of Dative Clitic Doubling in Bulgarian Ditransitives : An Applicative Analysis. *Journal of Slavic Linguistics*, *16*(1), 139–166. http://doi.org/10.1353/jsl.0.0012

Slioussar, N. (2011). Processing of a free word order language: the role of syntax and context. *Journal of Psycholinguistics Research*, *40*(4), 291–306. http://doi.org/10.1007/s10936-011-9171-5

Smolka, E., Komlósi, S., & Rösler, F. (2009). When semantics means less than morphology: The processing of German prefixed verbs. *Language and Cognitive Processes*, *24*(3), 337–375. http://doi.org/10.1080/01690960802075497

Sorace, A. (2007). Optionality at the syntax-discourse interface in near-native L2 speakers. *Second Language*, *6*, 3–15. Retrieved from http://ci.nii.ac.jp/naid/40015822591/

Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, *1*, 1–33. http://doi.org/10.1075/lab.1.1.01sor

Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, *13*(339810), 195–210. http://doi.org/10.1177/1367006909339810

Staron, M., & Kail, M. (2004). Off-line sentence comprehension in Polish monolinguals: developmental study based on the Competition Model. *Psychology of Language and Communication*, *8*(1), 63–70.

Stiebels, B., & Wunderlich, D. (1994). Morphology feeds syntax: the case of particle verbs. *Linguistics*, *32*(6), 913–968. http://doi.org/10.1515/ling.1994.32.6.913

Stoops, A., Luke, S. G., & Christianson, K. (2014). Animacy information outweighs morphological cues in Russian. *Language, Cognition and Neuroscience*, *29*(5), 584–604. http://doi.org/10.1080/01690965.2013.813560

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, *105*(2), 477–488. http://doi.org/10.1016/j.cognition.2006.10.009

Svenonius, P. (1996a). The Optionality of Particle Shift. *Working Papers in Scandinavian Syntax*, *57*(1), 47–75.

Svenonius, P. (1996b). *The Verb-Particle Alternation in the Scandinavian Languages*. Retrieved from http://norms.uit.no/thematic/workshops/scanvprtholeikanger.pdf

Taraldsen, K. T. (1983). *Parametric variation in phrase structure: A case study*. Universitet i Tromsö.
Taraldsen, K. T. (2000). V-movement and VP-movement in Derivations Leading to VO order. In P. Svenonius (Ed.), *The Derivation of VO and OV* (pp. 97–122). Amsterdam: John Benjamins Publishing Company.
Tomlin, R. S. (1986). *Basic word order: functional principles*. London: Croom Helm.
Traxler, M. J. (2005). Plausibility and Verb Subcategorization in Temporarily Ambiguous Sentences: Evidence from Self-Paced Reading. *Journal of Psycholinguistic Research*, *34*(1), 1–30. http://doi.org/10.1007/s10936-005-3629-2
Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing Subject and Object Relative Clauses: Evidence from Eye Movements. *Journal of Memory and Language*, *47*(1), 69–90. http://doi.org/10.1006/jmla.2001.2836
Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454–475. http://doi.org/10.1006/jmla.1996.0025
Tsimpli, I.-M., & Sorace, A. (2006). Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. In *Proceedings of the Annual Boston University Conference on Language Development* (Vol. 54, pp. 653–664). Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:No+Title#0
Ullman, M. T. (2001a). A neurocognitive perspective on language: the declarative/procedural model. *Nature Reviews. Neuroscience*, *2*(10), 717–26. http://doi.org/10.1038/35094573
Ullman, M. T. (2001b). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, *30*(1), 37–69. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11291183
Ullman, M. T. (2001c). The neural basis of lexicon and grammar in first and second language : the declarative / procedural model. *Bilingualism : Language and Cognition*, *4*(2), 105–122. http://doi.org/10.1017/S1366728901000220
Uszkoreit, H. (1986). Constraints on order. *Linguistics*, *24*(5), 883–906. http://doi.org/10.1515/ling.1986.24.5.883
Uszkoreit, H., Brants, T., Duchier, D., Krenn, B., Konieczny, L., Oepen, S., & Skut, W. (1998). Studien zur performanzorientierten Linguistik. *Kognitionswissenschaft*, *7*(3), 129–133. http://doi.org/10.1007/s001970050065
van Dommelen, W. A., & Husby, O. A. (2007). The Perception of Norwegian Word Tones by Chinese and German Listeners. In *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech* (pp. 165–174).
VanPatten, B., Keating, G. D., & Leeser, M. J. (2012). Missing verbal inflections as a representational problem: Evidence from self-paced reading. *Linguistic Approaches to Bilingualism*, *2*(2), 109–140. http://doi.org/10.1075/lab.2.2.01pat
Verhoeven, E. (2015). Thematic Asymmetries Do Matter! A Corpus Study of German Word Order. *Journal of Germanic Linguistics*, *27*(1), 45–104. http://doi.org/10.1017/S147054271400021X
Vikner, S. (1994). Scandinavian object shift and West Germanic scrambling. In N.

Corver & H. van Riemsdijk (Eds.), *Studies on scrambling* (pp. 487–517).
Vikner, S. (2006). Object Shift. In M. Everaert (Ed.), *The Blackwell Companion to Syntax* (pp. 392–436). Malden, Massachussetts: Blackwell.
Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*(1), 79–112. http://doi.org/10.1016/S0010-0277(02)00087-2
Webelhuth, G. (1989). *Syntactic saturation phenomena and the Germanic languages*. University of Massachusetts.
Weber, A., & Müller, K. (2004). Word order variation in German main clauses : A corpus analysis. In S. Hansen-Schirra, S. Oepen, & H. Uszkoreit (Eds.), *The 20th international conference on computational linguistics. Proceedings* (pp. 71–77). Geneva.
Weinberg, A. (1993). Parameters in the theory of sentence processing: Minimal Commitment theory goes east. *Journal of Psycholinguistic Research*, *22*(3), 339–364. http://doi.org/10.1007/BF01068016
White, L. (2003). Fossilization in steady state L2 grammars : Persistent problems with inflectional morphology. *Bilingualism : Language and Cognition*, *6*(2), 129–141. http://doi.org/10.1017/S1366728903001081
Williams, J. N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, *22*, 509–540. http://doi.org/10.1017/S0142716401004027
Witzel, N., Witzel, J., & Forster, K. I. (2012). Comparisons of online reading paradigms: eye tracking, moving-window, and maze. *Journal of Psycholinguistic Research*, *41*(2), 105–28. http://doi.org/10.1007/s10936-011-9179-x
Woolford, E. (2006). Lexical Case, Inherent Case, and Argument Structure. *Linguistic Inquiry*, *37*(1), 111–130.
Wurmbrand, S. (2000). The structure(s) of particle verbs. *Unpublished Manuscript*. Retrieved from http://wurmbrand.uconn.edu/Papers/structure-particles.pdf
Year, J. E. (2003). Sentence Processing within the Competition Model. *Columbia University Working Papers in TESOL & Applied Linguistics*, *3*, 1–27.
Zeevat, H., & Jäger, G. (2002). A Reinterpretation of Syntactic Alignment. In D. de Jongh, M. Nilsenová, & H. Zeevat (Eds.), *Proceedings of the 3rd and 4th International Symposium on Language, Logic and Computation* (pp. 1–15). Amsterdam.
Zeller, J. (2002). Particle verbs are heads and phrases. In N. Dehé, R. Jackendoff, A. McIntyre, & S. Urban (Eds.), *Verb-Particle Explorations* (pp. 233–267). Berlin: Mouton de Gruyter.
Zervakis, J., & Mazuka, R. (2013). Effect of Repeated Evaluation and Repeated Exposure on Acceptability Ratings of Sentences. *Journal of Psycholinguistic Research*, *42*(6), 505–525. http://doi.org/10.1007/s10936-012-9233-3

APPENDIX A – Materials

Study 1: Norwegian object topicalization[18]

Pilot study

Experimental items

	Inanimate subject	**Animate subject**
02	Traktoren vil dra oksen. ’The tractor will draw the ox.’	Oksen vil dra traktoren. ’The ox will draw the tractor.’
04	Ballen vil treffe spilleren. ’The ball will hit the player.’	Spilleren vil treffe ballen. ’The player will hit the ball.’
05	Bilen vil ramme elgen. ’The car will ram the elk.’	Elgen vil ramme bilen. ’The elk will ram the car.’
06	Døren vil berøre gutten. ’The door will touch the boy.’	Gutten vil berøre døren. ’The boy will touch the door.’
08	Fjæren vil streife jenten. ’The feather will touch the girl.’	Jenten vil streife fjæren. ’The girl will touch the feather.’
09	Ambulansen vil passere legen. ’The ambulance will pass the doctor.’	Legen vil passere ambulansen. ’The doctor will pass the ambulance.’
11	Sengeteppet vil gjemme småbarnet. ’The blanket will hide the toddler.’	Småbarnet vil gjemme sengeteppet. ’The toddler will hide the blanket.’
12	Kosten vil skyve katten. ’The broom push the cat.’	Katten vil skyve kosten. ’The cat will push the broom.’
13	Kanoen vil bære roeren. ’The canoe will carry the rower.’	Roeren vil bære kanoen. ’The rower will carry the canoe.’
14	Håndkleet vil tørke svømmeren. ’The towel will dry the swimmer.’	Svømmeren vil tørke håndkleet. ’The swimmer will dry the towel.’
15	Toget vil velte demonstranten. ’The train will knock over the protestor.’	Demonstranten vil velte toget. ’The protestor wll knock over the train.’
16	Opprøret vil styrte diktatoren. ’The uprising will overthrow the dictator.’	Diktatoren vil styrte opprøret. ’The dictator will overthrow the uprising.’
17	Annonsen vil søke vinneren. ’The ad will search the winner.’	Vinneren vil søke annonsen. ’The winner will search the ad.’
18	Historien vil forandre politikeren. ’The story will change the politician.’	Politikeren vil forandre historien. ’The politician will change the story.’
19	Fiskegarnet vil omgi dykkeren. ’The fishing net will surround the diver.’	Dykkeren vil omgi fiskegarnet. ’The diver will surround the fishing net.’
20	Steinen vil skubbe klatreren. ’The stone will push the climber.’	Klatreren vil skubbe steinen. ’The climber will push the stone.’
21	Båten vil taue fiskeren. ’The boat will tow the fisherman.’	Fiskeren vil taue båten. ‘The fisherman will tow the boat.’
22	Fyrtårnet vil belyse kapteinen. ’The lighthouse will illuminate the	Kapteinen vil belyse fyrtårnet. ’The captain will illuminate the

[18]NB! The numbering of items from the experiments on Norwegian remains constant from the pretest until the online or offline experiment. In this way different versions of the same item can be identified more easily across tasks. In the German experiments those items repeating the same critical verb receive the same item number across tasks.

	captatin.’	lighthouse.’
23	Datamaskinen vil beseire mesteren. ’The computer will defeat the champion.’	Mesteren vil beseire datamaskinen. ’The champion will defeat the computer.’
24	Teppet vil klappe spedbarnet. ’The carpet will stroke the baby.’	Spedbarnet vil klappe teppet. ‘The baby will stroke the carpet.’
25	Kameraet vil passere løperen. ’The camera will pass the runner.’	Løperen vil passere kameraet. ’The runner will pass the camera.’
26	Seilbrettet vil bære eventyreren. ’The surfboard will carry the adventurer.’	Eventyreren vil bære seilbrettet. ’The adventurer will carry the surfboard.’
27	Raset vil miste skiløperen. ’The avalanche will miss the skier.’	Skiløperen vil miste raset. ’The avalanche will miss the skier.’
28	Motivet vil forråde tyven. ’The motive will give away the thief.’	Tyven vil forråde motivet. ’The thief will give away the motive.’
29	Tornadoen vil forfølge forskeren. ’The tornado will chase the researcher.’	Forskeren vil forfølge tornadoen. ‘The researcher will chase the tornado.’
31	Bussen vil miste studenten. ’The bus will miss the student.’	Studenten vil miste bussen. ’The student will miss the bus.’
32	Kranen vil skyve arbeideren. ’The crane will push the worker.’	Arbeideren vil skyve kranen. ’The worker will push the crane.’
33	Skjulet vil forråde røveren. ’The hideout will give away the robber.’	Røveren vil forråde skjulet. ’The robber will give away the hideout.’
34	Planken vil treffe hopperen. ’The board will hit the jumper.’	Hopperen vil treffe planken. ’The jumper will hit the board.’
35	Gjerdet vil velte hesten. ’The fence will know down the horse.’	Hesten vil velte gjerdet. ’The horse will knock down the fence.’
36	Protesen vil forandre atleten. ‘The prosthesis will change the athlete.’	Atleten vil forandre protesen. ‘The athlete will change the prosthesis.’
37	Skogen vil omringe aktivistene. ’The wood will surround the activists.’	Aktivistene vil omringe skogen. ‘The activists will surround the wood.’
38	Ringene vil holde akrobaten. ’The rings will hold the acrobat.’	Akrobaten vil holde ringene. ‘The acrobat will hold the rings.’
39	Bomben vil sprenge eksperten. ’The bomb will blow up the expert.’	Eksperten vil sprenge bomben. ‘The expert will blow up the bomb.’
40	Flagget vil berøre kongen. ’The flag will touch the king.’	Kongen vil berøre flagget. ’The king will touch the flag.’
41	Konserten vil forstyrre naboen. ’The concert will disturb the neighbor.’	Naboen vil forstyrre konserten. ’The neighbor will disturb the concert.’
42	Radaren vil lokalisere soldaten. ’The radar will localize the soldier.’	Soldaten vil lokalisere radaren. ’The soldier will localize the radar.’

Filler items

01	Mannen vil overraske brevet. ’The man will surprize the letter.’
03	Treet vil kysse kvinnen. ’The tree will kiss the woman.’
07	Brødet vil spise hunden. ’The bread will eat the dog.’
10	Dronningen vil kjøre huset. ’The queen will drive the house.’
30	Elven vil drikke moren. ’The river will drink the mother.’
43	Hamsteren vil bake kaken.

	'The hamster will bake the cake.'
44	Håret vil klippe eleven. 'The hair will cut the pupil.'
45	Læreren vil lese verden. 'The teacher will read the world.'
46	Skoen vil binde enken. 'The shoe will bind the window.'
47	Bussjaføren vil plukke flyet. 'The busdriver will pick the airplane.'
48	Flasken vil synge sauen. 'The bottle will sing the sheep.'
49	Rørleggeren vil installere klaveret. 'The plumber will install the piano.'

Experiment 1a - Agent identification task

Experimental items

	SVO	OVS
08	Fjæren skulle streife jenten. 'The feather should touch the girl.'	Jenten skulle fjæren streife. 'The girl, the feather should touch.'
13	Roeren ville bære kanoen. 'The rower wanted to carry the canoe.'	Kanoen ville roeren bære. 'The canoe, the rower wanted to carry.'
19	Fiskegarnet vil omgi dykkeren. 'The fishing net will surround the diver.'	Dykkeren vil fiskegarnet omgi. ‚The diver, the fishing net will surround.'
24	Spedbarnet kan klappe teppet. ‚The baby can stroke the carpet.'	Teppet kan spedbarnet klappe. 'The carpet, the baby can stroke.'
31	Studenten skal miste bussen. 'The student shall miss the bus.'	Bussen skal studenten miste. 'The bus, the student shall miss.'
35	Hesten kunne velte gjerdet. 'The horse could knock down the fence.'	Gjerdet kunne hesten velte. ‚The fence, the horse could knock down.'
50	Rørleggeren må hjelpe læreren. ‚The plumber has to help the teacher.'	Læreren må rørleggeren hjelpe. 'The teacher, the plumber has to help.'
51	Bussjaføren måtte kysse dronningen. 'The bus driver had to kiss the queen.'	Dronningen måtte bussjaføren kysse. 'The queen, the bus driver had to kiss.'
52	Postbudet måtte belære sjefen. ‚The postman had to instruct the boss.'	Sjefen måtte postbudet belære. 'The boss, the postman had to instruct.'
53	Prinsen ville male prinsessen. 'The prince wanted to paint the princess.'	Prinsessen ville prinsen male. 'The princess, the prince wanted to paint.'
54	Løven skal drepe jegeren. 'The lion shall kill the hunter.'	Jegeren skal løven drepe. ‚The hunter, the lion shall kill.'
55	Brannmannen må redde politimannen. ‚The firefighter has to rescue the policeman.'	Politimannen må brannmannen redde. 'The policeman, the firefighter has to rescue.'

Filler items

01	Mannen overrasker brevet. ‚The man, the letter surprizes.'
56	Kuen melker bonden. ‚The cow, the farmer milks.'
57	Steinen knuser vinduet. 'The stone breaks the window.'

58	Eleven mater sauen. 'The pupil feeds the sheep.'
59	Det er vandreren som kunne fotografere indianeren. 'It is the hiker who could take a picture of the Native American.'
60	Det er herren som hunden vil savne. 'It is the master the dog will miss.'
61	Det er nordmannen som skulle kjenne japaneren. 'It is the Norwegian who should know the Japanese.'
62	Det er amerikaneren som tyskeren kan elske. 'It is the American who the German can love.'
63	Det er sveitseren som ville beseire russen. 'It is the Swiss who wanted to beat the Russian.'
64	Det er svenskeren som dansken måtte forlate. 'It is the Swede who the Dane had to leave.'
65	Det er danseren som må underholde forretningsmannen. 'It is the dancer who has to entertain the businessman.'
66	Det er moren som barnet skal tape. 'It is the mother who the child shall lose.'
67	På mandag slår dommeren bokseren. 'On Monday, the referee beats the boxer.'
68	Siste juli giftet musen kunstneren. 'Last July, the muse married the artist.'
69	Neste uke møter læreren foreldrene. 'Next week, the teacher meets the parents.'
70	Om sommeren jager reven ulven. 'In Summer, the fox hunts the wolf.'

Experiment 1b - SPR task

Experimental items

	SVO_ inanimate	**OVS_ inanimate**
	SVO_animate	**OVS_animate**
02	Traktoren som er litt problematisk, vil dra oksen til bondegården. 'The tractor that is a bit problematic, will pull the ox to the farm.'	Oksen som er litt problematik, vil traktoren dra til bondegården. 'The ox that is a bit problematic, the tractor will draw to the farm.'
	Oksen som er litt problematik, vil dra traktoren til bondegården. 'The ox that is a bit problematic, will pull the tractor to the farm.'	Traktoren som er litt problematisk, vil oksen dra til bondegården. 'The tractor that is a bit problematic, the tractor will draw to the farm.'
04	Ballen som er ennå ny, vil treffe spilleren foran målet. 'The ball that is still new, will hit the player in front of the goal.'	Spilleren som er ennå ny, vil ballen treffe foran målet. 'The player who is still new, the ball will hit in front of the goal.'
	Spilleren som er ennå ny, vil treffe ballen foran målet. 'The player who is still new, will hit the ball in front of the goal.'	Ballen som er ennå ny, vil spilleren treffe foran målet. 'The ball that is still new, the player will hit in front of the goal.'
05	Bilen som ikke er liten, vil ramme elgen i skogen. 'The car that is not small, will hit the elg in the woods.'	Elgen som ikke er liten, vil bilen ramme i skogen. 'The elg that is not small, the car will hit in the woods.'
	Elgen som ikke er liten, vil ramme bilen i skogen.	Bilen som ikke er liten, vil elgen ramme i skogen.

	’The elg that is not small, will hit the car in the woods.’	’The car that is not small, the elg will hit in the woods.’
06	Døren som allerede er åpen, vil berøre gutten i morgen. ’The door that is already open, will touch the boy in the morning.’	Gutten some allerede er beruset, vil døren berøre i morgen. ’The boy who is already drunk, the door will touch in the morning.’
	Gutten some allerede er beruset, vil berøre døren i morgen. ’The boy who is already drunk, will touch the door in the morning.’	Døren som allerede er åpen, vil gutten berøre i morgen. ’The door who is already open, the boy will touch in the morning.’
09	Ambulansen som alltid er punktlig, vil passere legen bak sykehuset. ’The ambulance that is always on time, will pass the doctor behind the hospital.’	Legen som alltid er punktlig, vil ambulansen passere bak sykehuset. ’The doctor who is always on time, the ambulance will pass behind the hospital.’
	Legen som alltid er punktlig, vil passere ambulansen bak sykehuset. ’The doctor who is always on time, will pass the ambulance behind the hospital.’	Ambulansen som alltid er punktlig, vil legen passere bak sykehuset. ’The ambulance that is always on time, the doctor will pass behind the hospital.
11	Sengeteppet som er veldig tjukt, vil gjemme småbarnet for tyven. ’The blanket that is very thick, will hide the toddler from the thief.’	Småbarnet som er veldig søtt, vil sengeteppet gjemme for tyven. ’The toddler who is very cute, the blanket will hide from the thief.’
	Småbarnet som er veldig søtt, vil gjemme sengeteppet for tyven. ‘The toddler who is very cute, will hide the blanket from the thief.’	Sengeteppet som er veldig tjukt, vil småbarnet gjemme for tyven. ’The blanket that is very thick, the toddler will hide from the thief.’
14	Håndkleet som sikkert er pent, vil tørke svømmeren på stranden. ’The towel that is surely pretty, will dry the swimmer on the beach.’	Svømmeren som sikkert er pen, vil håndkleet tørke på stranden. ’The swimmer who surely is pretty, the towel will dry on the beach.’
	Svømmeren som sikkert er pen, vil tørke håndkleet på stranden. ’The swimmer who is surely pretty, will dry the towel on the beach.’	Håndkleet som sikkert er pent, vil svømmeren tørke på stranden. ’The towel that surely is pretty, the swimmer will dry on the beach.’
16	Opprøret som faktisk er voldelig, vil styrte diktatoren i begynnelsen. ’The uprising that is actually violent, will overthrow the dictator in the beginning.’	Diktatoren som faktisk er voldelig, vil opprøret styrte i begynnelsen. ’The dictator who is actually violent, the uproar will overthrow in the beginning.’
	Diktatoren som faktisk er voldelig, vil styrte opprøret i begynnelsen. ’The dictator who is actually violent, will overthrow the uprising in the beginning.’	Opprøret som faktisk er voldelig, vil diktatoren styrte i begynnelsen. ’The uprising that is actually violent, the dictator will overthrow in the beginning.’
18	Historien som er veldig spennende, vil forandre politikeren for alltid. ’The story that is very exciting, will change the politician for good.’	Politikeren som er veldig populær, vil historien forandre for alltid. ‘The politician who is very popular, the story will change for good.’
	Politikeren som er veldig populær, vil forandre historien for alltid. ’The politician who is very popular, will change the story for good.’	Historien som er veldig spennende, vil politikeren forandre for alltid. ’The story who is very exciting, the politician will change for good.’
21	Båten som nesten er uskadd, vil taue fiskeren til havnen. ’The boat that is nearly undamaged, will	Fiskeren som nesten er uskadd, vil båten taue til havnen. ’The fisherman who is nearly unhurt,

	tow the fisherman to the harbor.'	the boat will tow to the harbor.'
	Fiskeren som nesten er uskadd, vil taue båten til havnen. 'The fisherman who is nearly unhurt, will tow the boat to the harbor.'	Båten som nesten er uskadd, vil fiskeren taue til havnen. 'The boat that is nearly undamaged, the fisherman will tow to the harbor.'
22	Fyrtårnet som ofte er forlatt, vil belyse kapteinen i stormen. 'The lighthouse that is often abandoned, will illuminate the captain in the storm.'	Kapteinen som ofte er ensome, vil fyrtårnet belyse i stormen. 'The captain who is often lonely, the lighthouse will illuminate in the storm.'
	Kapteinen som ofte er ensome, vil belyse fyrtårnet i stormen. 'The captain who is often lonely, will illuminate the lighthouse in the storm.'	Fyrtårnet som ofte er forlatt, vil kapteinen belyse i stormen. 'The lighthouse that is often abandoned, the captain will illuminate in the storm.'
23	Datamaskinen som er uvanlig avansert, vil beseire mesteren i sjakk. 'The computer that is unusually advanced, will defeat the champion in chess.'	Mesteren som er uvanlig avansert, vil datamaskinen beseire i sjakk. 'The champion who is unusually advanced, the computer will defeat in chess.'
	Mesteren som er uvanlig avansert, vil beseire datamaskinen i sjakk. 'The champion who is unusually advanced, will defeat the computer in chess.'	Datamaskinen som er uvanlig avansert, vil mesteren beseire i sjakk. 'The computer that is unusually advanced, the champion will defeat in chess.'
25	Kameraet som er ganske raskt, vil passere løperen ved start. 'The camera that is pretty fast, will pass the runner at the start.'	Løperen som er ganske rask, vil kameraet passere ved start. 'The runner who is pretty fast, the camera will pass at the start.'
	Løperen som er ganske rask, vil passere kameraet ved start. 'The runner who is pretty fast, will pass the camera at the start.'	Kameraet som er ganske raskt, vil løperen passere ved start. 'The camera that is pretty fast, the runner will pass at the start.'
26	Seilbrettet som er fryktelig tynt, vil bære eventyreren ned elven. 'The surfboard that is terribly thin, will carry the adventurer down the river.'	Eventyreren som er fryktelig tynn, vil seilbrettet bære ned elven. 'The adventurer who is terribly thin, the surfboard will carry down the river.'
	Eventyreren som er fryktelig tynn, vil bære seilbrettet ned elven. 'The adventurer who is terribly thin, will carry the surfboard down the river.'	Seilbrettet som er fryktelig tynt, vil eventyreren bære ned elven. 'The surfboard that is terribly thin, the adventurer will carry down the river.'
29	Tornadoen som kanskje er farlig, vil forfølge forskeren langs veien. 'The tornado that is maybe dangerous, will follow the scientist down the road.'	Forskeren som kanskje er gal, vil tornadoen forfølge langs veien. 'The scientist who is maybe crazy, the tornado will follow down the road.'
	Forskeren som kanskje er gal, vil forfølge tornadoen langs veien. 'The scientist who is maybe crazy, will follow the tornado down the road.'	Tornadoen som kanskje er farlig, vil forskeren forfølge langs veien. 'The tornado that is maybe dangerous, the scientist will follow down the road.'
32	Kranen some plutselig er ukontrollert, vil skyve arbeideren fra byggeplassen. 'The crane that is suddenly uncontrolled, will push the worker from the construction site.'	Arbeideren som plutselig er syk, vil kranen skyve fra byggeplassen. 'The worker who is suddenly sick, the crane will push from the construction site.'
	Arbeideren som plutselig er syk, vil skyve	Kranen some plutselig er ukontrollert,

	kranen fra byggeplassen. ‘The worker who is suddenly sick, will push the crane from the construction site.’	vil arbeideren skyve fra byggeplassen. ’The crane that is suddenly uncontrolled, the worker will push from the construction site.’
34	Planken som er litt gammel, vil treffe hopperen før finalen. ‘The board that is a bit old, will hit the jumper before the final.’	Hopperen som er litt gammel, vil planken treffe før finalen. ’The jumper who is a bit old, the board will hit before the final.’
	Hopperen som er litt gammel, vil treffe planken før finalen. ’The jumper who is a bit old, will hit the board before the final.’	Planken som er litt gammel, vil hopperen treffe før finalen. ‘The board that is a bit old, the jumper will hit before the final.’
36	Protesen som er enormt heldig, vil forandre atleten etter konkurransen. ’The prosthesis that is enormously successful, will change the athlete after the competition.’	Atleten som er enormt heldig, vil protesen forandre etter konkurransen. ’The athlete who is enormously successful, the prosthesis will change after the competition.’
	Atleten som er enormt heldig, vil forandre protesen etter konkurransen. ’The athlete who is enormously successful, will change the prosthesis after the competition.’	Protesen som er enormt heldig, vil atleten forandre etter konkurransen. ’The prosthesis that is enormously successful, the athlete will change after the competition.’
37	Skogen som åpenbart er stor, vil omringe aktivistene som beskyttelse. ‘The wood that is obviously big, will surround the activists as protection.’	Aktivistene som åpenbart er motiverte, vil skogen omringe som beskyttelse. ‘The activists that are obviously motivated, the wood will surround as protection.’
	Aktivistene som åpenbart er motiverte, vil omringe skogen som beskyttelse. ‘The activists that are obviously motivated, will surround the wood as protection.’	Skogen som åpenbart er stor, vil aktivistene omringe som beskyttelse. ‘The wood that is obviously big, the activists will surround as protection.’
38	Ringene som aldri er brukte, vil holde akrobaten i luften. ‘The rings that are never used, will hold the acrobat in the air.’	Akrobaten som aldri er uforberedt, vil ringene holde i luften. ‘The acrobat that is never unprepared, the rings will hold into the air.’
	Akrobaten som aldri er uforberedt, vil holde ringene i luften. ‘The acrobat that is never unprepared, will hold the rings into the air.’	Ringene som aldri er brukte, vil akrobaten holde i luften. ‘The rings that are never used, the acrobat will hold in the air.’
39	Bomben som er utrolig rusten, vil sprenge eksperten i stykker. ‘The bomb that is incredibly rusty, will blow the expert into pieces.’	Eksperten som er utrolig modg, vil bomben sprenge i stykker. ‘The expert that is incredibly brave, the bomb will blow into pieces.’
	Eksperten som er utrolig modg, vil sprenge bomben i stykker. ‘The expert that is incredibly brave, will blow the bomb into pieces.’	Bomben som er utrolig rusten, vil eksperten sprenge i stykker. ‘The bomb that is incredibly rusty, the expert will blow into pieces.’
40	Flagget som dessverre er vått, vil berøre kongen på nasjonaldagen. ‘The flag that is unfortunately wet, will touch the king on the national day.’	Kongen som dessverre er våt, vil flagget berøre på nasjonaldagen. ’The king who is unfortunately wet, the flag will touch on the national day.’
	Kongen som dessverre er våt, vil berøre flagget på nasjonaldagen. ’The king who is unfortunately wet, will	Flagget som dessverre er vått, vil kongen berøre på nasjonaldagen. ‘The flag that is unfortunately wet, the

	touch the flag on the national day.’	king will touch on the national day.’
41	Konserten som er altfor høy, vil forstyrre naboen om natten. ’The concert that is way too loud, will disturb the neighbor at night.’	Naboen som er altfor følsom, vil konserten forstyrre om natten. ’The neighbor who is way too sensitive, forstyrre will disturb x at night.’
	Naboen som er altfor følsom, vil forstyrre konserten om natten. ’The neighbor who is way too sensitive, will disturb the concert at night.’	Konserten som er altfor høy, vil naboen forstyrre om natten. ’The concert that is way too loud, the neighbor will disturb at night.’
42	Radaren som vanligvis er korrekt, vil lokalisere soldaten i mørket. ’The radar that normally is correct, will localize the soldier in the darkness.’	Soldaten som vanligvis er korrekt, vil radaren lokalisere i mørket. ’The soldier who normally is correct, the radar will localize in the darkness.’
	Soldaten som vanligvis er korrekt, vil lokalisere radaren i mørket. ’The soldier who normally is correct, will localize the radar in the darkness.’	Radaren som vanligvis er korrekt, vil soldaten lokalisere i mørket. ’The radar that normally is correct, the soldier will localize in the darkness.’

Comprehension questions

	Inanimate subject	**Animate subject**	**Answer**	**L1 accuracy**	**L2 accuracy**
04	Treffer ballen spilleren? ’Does the ball hit the player?’	Treffer spilleren ballen? ‘Does the player hit the ball?’	Yes	68.75%	71.88%
09	Passeres ambulansen bak sykehuset? ‘Is the ambulance being passed behind the hospital?’	Passeres legen bak sykehuset? ‘Is the doctor being passed behing the hospital?’	No	43.75%	31.25%
18	Forandres politikeren for alltid? ’Is the politician being changed for good?’	Forandres historien for alltid? ‘Is the story changed for good?’	Yes	59.38%	71.88%
23	Slår mesteren datamaskinen? ’Does the champion beat the computer?’	Slår datamaskinen mesteren? ‘Does the computer beat the champion?’	No	71.88%	65.63%
36	Forandrer protesen atleten? ‘Does the prosthesis change the athlete?’	Forandrer atleten protesen? ‘Does the athlete change the prosthesis?’	Yes	65.63%	84.38%
38	Holdes ringene i luften? ‘Are the rings being held in the air?’	Holdes akrobaten i luften? ‘Is the acrobat being held in the air?’	No	53.13%	56.25%
41	Forstyrres naboen om natten? ’Is the neighbor being disturbed at night?’	Forstyrres konserten om natten? ‘Is the concert being disturbed at night?’	Yes	71.88%	84.38%
42	Lokaliserer soldaten raderen?	Lokaliserer radaren soldaten?	No	75%	62.5%

	‘Does the soldier localize the radar?	‘Does the radar localize the soldier?’			

Practice items and filler items

P1	Det driver en kano på vannet. ’There floats a canoe on the lake.’
P2	Roeren som er meget vennlig, vil hjelpe dykkeren med utstyret. ’The rower who is very friendly, will help the diver with the equipment.’
P3	De ser hverandre hver dag på arbeidet. ’They see each other every day at work.’
P4	Rune samler frimerker og Joakim samler dem også. ’Rune collects stamps and Joakim collects them as well.’
71	Kristoffer drikker gjerne øl, men Ida drikker det ikke. ’Kristoffer likes to drink beer, but Ida does not drink it.’
72	Katrine vet ikke svaret, men Tor vet det. ’Katrine doesn’t know the answer, but tor knows it.’
73	Anette liker ikke bildet og Per liker det ikker heller. ’Anette doesn’t like the picture and Per does not like it either.’
74	Vanligvis gjør ikke Geir oppvasken, men han gjør den i dag. ’Usually, Geir does not do the dishes, but he does them today.’
75	Om morgenen vanner Liv aldri blomstene, men hun gjorde det i morges. ’In the mornings, Liv never waters the flowers, but she did it this morning.’
76	Vanligvis er Gunn kjempesterk, men i går vant hun ikke løpet. ’Usually, Gunn is very strong, but yesterday she did not win the race.’
77	Om kvelden er Ragnhild ofte trøtt, men i kveld er hun våken. ’In the evening, Ragnhild is often tired, but tonight she is awake.’
78	Jenten som er veldig lykkelig, vil klappe hesten på ryggen. ’The girl who is very happy, will pat the horse on the back.’
79	Demonstranten som sjelden er uinvolvert, vil fortelle studenten om naturen. ’The protestor who rarely is uninvolved, will tell the student about nature.
80	Tyven some er utrolig raffinert, vil bestjele røveren uten problemer. ’The thief who is unbelievably clever, will steal from the robber without problems.
81	Raset som er enormt raskt, vil begrave huset under snøen. ’The avalanche that is enormously fast, will bury the house under the snow.
82	Bussen som er veldig koselig, vil krysse elven med fergen. ’The bus that is very cosy, will cross the river by ferry.’
83	Kosten som er svært gammel, vil feie hårene under teppet. ’The broom that is pretty old, will sweep the hairs under the carpet.’
84	Toget som alltid er sent, vil transportere trærne til kysten. ’The train that is always late, will transport the trees to the coast.’
85	Hvis vi hadde tid, ville vi skrive brev til bestemor. ’If we had time, we would write a letter to grandmother.’
86	Hvis han hadde penger, ville han reise omkring i verden. ’If he had money, he would travel around the world.’
87	Han forklarer at hun bryr seg om finansene våre. ’He explains that she takes care of our finances.’
88	Vi observerer at de ser seg i speilet. ’We observe that they see themselves in the mirror.’
89	I går ble det kastet en stein på vinduet. ’Yesterday a stone was thrown through the window.’
90	Det finnes mange annonser i denne avisen. ’There are many ads in this newspaper.’
91	Dere trives ikka så bra i Oslo. ’You are not enjoying yourself in Oslo.’

92	Vi skal møtes på kurset om en uke. ’We shall meet in the course in one week.’
93	Hun synes at hybelen hennes er ganske dyr. ’She thinks her (own) apartment is pretty expensive.’
94	Han glemte å mate kattene hans i ett døgn. ’He forgot to feed his (another 3rd person) cats one day long.’

Comprehension questions for practice and filler items

	Question	**Answer**	**L1 accuracy**	**L2 accuracy**
P4	Samler Joakim frimerker? ’Does Joakim collect stamps?’	Yes	-	-
72	Vet Tor svaret? ‘Does Tor know the answer?’	Yes	100%	96.88%
74	Gjør Geir oppvasken alle dager? ’Does Geir do the dishes every day?’	No	96.88%	100%
80	Vil røveren bestjele tyven? ’Will the robber steal from the thief?’	No	37.5%	65.63%
81	Ligger huset begravet under snøen? ’Does the house lie buried under the snow?’	Yes	25%	34.38%
84	Er toget alltid sent? ’Is the train always late?’	Yes	96.88%	87.5%
87	Bryr han seg om finansene? ’Does he take care of the finances?’	No	31.25%	31.25%
89	Er vinduet slått i stykker? ’Is the window broken into pieces?’	Yes	18.75%	40.63%
94	Glemte han a måte kattene sine? ’Did he forget to feed his own cats?’	No	12.5%	21.88%

Study 2: German ditransitive sentences

Experiment 2a - Acceptability rating task

Experimental items

	Dative-first	Accusative-first
01	Der Abteilungsleiter hat dem Kollegen den Zettel zugesteckt.	Der Abteilungsleiter hat den Zettel dem Kollegen zugesteckt.
	‘The manager slipped the colleague the note.’	
02	Der Sänger hat dem Publikum den Song vorgesungen.	Der Sänger hat den Song dem Publikum vorgesungen.
	‘The singer sang the song to the audience.’	
03	Der Schauspieler hat dem Regisseur den Monolog vorgespielt.	Der Schauspieler hat den Monolog dem Regisseur vorgespielt.
	‘The actor played the monologue for the director.’	
04	Der Schriftsteller hat dem Ausschuss den Aufsatz vorgetragen.	Der Schriftsteller hat den Aufsatz dem Ausschuss vorgetragen.
	‘The writer recited the essay for the committee.’	
05	Der Assistent hat dem Professor den Vortrag ausgeredet.	Der Assistent hat den Vortrag dem Professor ausgeredet.
	‘The assistant talked the professor out of the lecture.’	
06	Die Chefin hat dem Kollegium den Mitarbeiter vorgestellt.	Die Chefin hat den Mitarbeiter dem Kollegium vorgestellt.
	‘The boss introduced the coworker to the college.’	
07	Die Enkelin hat dem Großvater den Kuchen mitgebracht.	Die Enkelin hat den Kuchen dem Großvater mitgebracht.
	‘The granddaughter brought cake for the grandfather.’	
08	Der Bräutigam hat dem Trauzeugen den Champagner ausgegeben.	Der Bräutigam hat den Champagner dem Trauzeugen ausgegeben.
	‘The groom bought the best man the champagne.’	

Filler items

09	Der Jäger hat die Wildschweine verjagt. ‘The hunter has chased away the wild boars.’
10	Der Informatiker hat die unlogischen Regeln missachtet. ‘The computer scientist has violated the illogical rules.’
11	Der Künstler hat die weiße Leinwand bemalt. ‘The artist has painted the white canvas.’
12	Der Passagier hat die Fahrkarte entwertet. ‘The passenger has punched the ticket.’
13	*Der Nussknacker zerkleinert die Nüsse aus. ‘The nutcracker crushes the nuts out.’
14	*Der Schüler versteht die Fremdsprache ein. ‘The student understands the foreign language in.’
15	*Der Diktator missbraucht seine Macht auf. ‘The dictator abuses his power up.’
16	*Der Gast bestellt das Abendessen an. ‘The guest orders lunch on.’
17	*Der Modedesigner entwirft die neue Kollektion weg.’ ‘The fashion designer designs the new collection away.’

18	*Der Vater zerreißt die Matheklausur ab. 'The father tears the math exam off.'
19	*Der Junge hat die Vokabeln wiedergeholt. 'The boy has gotten the vocabulary back.'
20	*Der Computer hat die Uhr umstellt. 'The computer has enclosed the clock.'
21	*Die Frau hat ihr Konto übergezogen. 'The woman has slipped her account on.'
22	*Der Fährmann hat die Fähre übersetzt. 'The ferryman has translated the ferry.'
23	*Die Journalistin hat den Begriff umgeschrieben. 'The journalist has rewritten the expression.'
24	*Der Kurier hat sein Fahrrad unterstellt. 'The messenger has insinuated his bicycle.'
25	*Die Regierung hat das neue Gesetz durchsetzt. 'The government has permeated the new law.'
26	*Der Schwimmer hat den Ärmelkanal durchgeschwommen. 'The swimmer has kept on swimming through the Channel.'

Experiment 2b - SPR task

Experimental items

<table>
<tr><td rowspan="2"></td><td colspan="2">Contextualizing sentences</td></tr>
<tr><td>Dative-first</td><td>Accusative-first</td></tr>
<tr><td rowspan="3">27</td><td colspan="2">Endlich sollte die große Villa verkauft werden. Der Kaufvertrag musste nur noch unterschrieben werden.
'Finally the big villa was supposed to be sold. The sales contract only needed to be signed.'</td></tr>
<tr><td>Die Sekretärin wusste, dass der Makler dem Ehepaar den Kaufvertrag geemailt hatte.</td><td>Die Sekretärin wusste, dass der Makler den Kaufvertrag dem Ehepaar geemailt hatte.</td></tr>
<tr><td colspan="2">'The secretary knew that the estate agent had emailed the sales contract to the couple.'</td></tr>
<tr><td rowspan="3">28</td><td colspan="2">Seit dem Tod ihrer Mutter hatte Eva oft die Schule geschwänzt. Irgendwann kam Zuhause ein Brief von der Schule an.
'Since the death of her mother Eva had skipped school often. Eventually a letter from school arrived at home.'</td></tr>
<tr><td>Eva glaubte, dass der Lehrer dem Vater den Brief geschrieben hatte.</td><td>Eva glaubte, dass der Lehrer den Brief dem Vater geschrieben hatte.</td></tr>
<tr><td colspan="2">'Eva thought that the teacher had written the letter to the father.'</td></tr>
<tr><td rowspan="3">29</td><td colspan="2">Frau Schmitz war alt und lebte sehr bescheiden. Völlig unerwartet erbte sie letzten Monat einen großen Geldbetrag.
'Mrs Schmitz was old and lived very humbly. Quite unexpectedly she inherited a large amount of money.'</td></tr>
<tr><td>Man vermutete, dass die Witwe dem Zoo den Betrag spenden würde.</td><td>Man vermutete, dass die Witwe den Betrag dem Zoo spenden würde.</td></tr>
<tr><td colspan="2">'One assumed that the widow would donate the sum to the zoo.'</td></tr>
<tr><td rowspan="2">30</td><td colspan="2">Der Kulturverein hatte eine Lesung organisiert. Der neueste Roman eines berühmten Autors sollte vorgestellt werden.
'The cultural association had organized a reading. The newest novel of a famous author should be presented.'</td></tr>
<tr><td>Es war geplant, dass der Autor dem Publikum den Roman vorlesen würde.</td><td>Es war geplant, dass der Autor den Roman dem Publikum vorlesen würde.</td></tr>
</table>

<table>
<tr><td></td><td colspan="2">‘It was planned that the author would read the novel to the audience.’</td></tr>
<tr><td rowspan="3">31</td><td colspan="2">Günther und Doris hatten in der Tanzschule einen Walzer einstudiert. Sie hofften, als Statisten in einem Ballett mitwirken zu dürfen.
‘Günther and Doris had rehearsed a waltz at the dance school. They hoped to be allowed to participate in a ballet as extras.’</td></tr>
<tr><td>Alle fanden es mutig, dass das Paar dem Regisseur den Walzer vortanzen wollte.</td><td>Alle fanden es mutig, dass das Paar den Walzer dem Regisseur vortanzen wollte.</td></tr>
<tr><td colspan="2">‘Everyone thought it brave that the couple wanted to dance the waltz for the director.’</td></tr>
<tr><td>32</td><td colspan="2">Der Prozess nahm eine überraschende Wendung, als endlich der Bericht des Gutachters eintraf.
‘The trial took an unexpected turn when the report of the expert finally arrived.’</td></tr>
<tr><td rowspan="2"></td><td>Schnell erfuhren die Medien, dass der Richter dem Staatsanwalt den Bericht vorgetragen hatte.</td><td>Schnell erfuhren die Medien, dass der Richter den Bericht dem Staatsanwalt vorgetragen hatte.</td></tr>
<tr><td colspan="2">‘The media soon learned that the judged had reported the report to the attorney.’</td></tr>
<tr><td rowspan="3">33</td><td colspan="2">Das U-Boot hätte beinahe ein Fischerboot gerammt. Ein Befehl des Leutnants war offenbar ignoriert worden.
‘The submarine had almost rammed the fishing boat. An order of the lieutenant had apparently been ignored.’</td></tr>
<tr><td>Der Kapitän bestätigte, dass der Leutnant dem Matrosen den Befehl zugerufen hatte.</td><td>Der Kapitän bestätigte, dass der Leutnant den Befehl dem Matrosen zugerufen hatte.</td></tr>
<tr><td colspan="2">‘The captain confirmed that the lieutenant had shouted the order to the sailor.’</td></tr>
<tr><td rowspan="3">34</td><td colspan="2">Der Rechtsanwalt der Firma prüfte die Bedingungen für die geplante Übernahme. Es gab bereits einen Vertragsentwurf.
‘The attorney of the company tested the condition for the planned take-over. There was already a draft agreement.’</td></tr>
<tr><td>Die Sekretärin schlug vor, dass der Anwalt dem Investor den Entwurf emailen sollte.</td><td>Die Sekretärin schlug vor, dass der Anwalt den Entwurf dem Investor emailen sollte.</td></tr>
<tr><td colspan="2">‘The secretary suggested that the attorney should email the draft to the investor.’</td></tr>
<tr><td rowspan="3">35</td><td colspan="2">Der junge Forscher hatte den Antrag für das internationale Projekt schon komplett ausgearbeitet.
‘The young research had already entirely written up the proposal for the international project .’</td></tr>
<tr><td>Die Institutsleiterin fand es falsch, dass der Forscher dem Kollegen den Antrag gefaxt hatte.</td><td>Die Institutsleiterin fand es falsch, dass der Forscher den Antrag dem Kollegen gefaxt hatte.</td></tr>
<tr><td colspan="2">‘The head of the institute thought it wrong that the researcher had faxed the proposal to the colleague.’</td></tr>
<tr><td rowspan="3">36</td><td colspan="2">Herr Müller wusste immer alles besser. Heute hatte er einen Ratschlag zur Verbesserung des Steuersystems.
‘Mr. Müller always knows better. Today he had an advice for the improvement of the tax system.’</td></tr>
<tr><td>Seine Frau befürchtete, dass der Rentner dem Finanzamt den Ratschlag schreiben würde.</td><td>Seine Frau befürchtete, dass der Rentner den Ratschlag dem Finanzamt schreiben würde.</td></tr>
<tr><td colspan="2">‘His wife feared that the retiree would write the advice to the tax authorities.’</td></tr>
<tr><td>37</td><td colspan="2">Kurz vor ihrem Tod hatte Frau Ziegler noch einen Hauptgewinn im Lotto gemacht. Sie hatte nur ein einziges Kind.
‘Shortly before her death Mrs. Ziegler had won the jackpot in the lottery. She had only one child.’</td></tr>
</table>

<table>
<tr><td rowspan="2"></td><td>Die Lokalzeitung schrieb, dass der Sohn dem Tierschutzverein den Hauptgewinn gespendet hatte.</td><td>Die Lokalzeitung schrieb, dass der Sohn den Hauptgewinn dem Tierschutzverein gespendet hatte.</td></tr>
<tr><td colspan="2">‘The local news wrote that the son had donated the jackpot to the society for the prevention of cruelty to animals.’</td></tr>
<tr><td rowspan="3">38</td><td colspan="2">In der Zeitung stand gestern ein Artikel über ein neues Medikament gegen Krebs.
‘Yesterday there was an article about a new drug against cancer in the newspaper.’</td></tr>
<tr><td>Stefan fand es unverantwortlich, dass die Mutter dem Großvater den Artikel vorgelesen hatte.</td><td>Stefan fand es unverantwortlich, dass die Mutter den Artikel dem Großvater vorgelesen hatte.</td></tr>
<tr><td colspan="2">‘Stefan thought it irresponsible that the mother had read the article to the grandfather.’</td></tr>
<tr><td rowspan="3">39</td><td colspan="2">Erikas Onkel hatte mal wieder einen neuen Song komponiert. Er hielt sich für einen genialen Musiker.
‘Erika’s uncle had once again composed a new song. He considered himself an ingenious musician.’</td></tr>
<tr><td>Erika war entsetzt, dass der Onkel dem Chorleiter den Song vorgesungen hatte.</td><td>Erika war entsetzt, dass der Onkel den Song dem Chorleiter vorgesungen hatte.</td></tr>
<tr><td colspan="2">‘Erika was appalled that the uncle had sung the song to the choir director.’</td></tr>
<tr><td rowspan="3">40</td><td colspan="2">Trotz seiner Krankheit hatte der alte Regisseur noch einen letzten Spielfilm gedreht.
‘Despite his sickness the old director had shot one last feature film.</td></tr>
<tr><td>Der Produzent war erfreut, dass der Künstler dem Kritiker den Spielfilm vorspielen durfte.</td><td>Der Produzent war erfreut, dass der Künstler den Spielfilm dem Kritiker vorspielen durfte.</td></tr>
<tr><td colspan="2">‘The producer was delighted that the artist was allowed to play the feature film to the critic.’</td></tr>
<tr><td rowspan="3">41</td><td colspan="2">Das junge Paar wollte ganz traditionell heiraten. Immer wieder übten sie einen bestimmten Tanzschritt.
‘The young couple wanted to marry traditionally. The practiced one particular dancing step again and again.’</td></tr>
<tr><td>Der Bräutigam ahnte nicht, dass die Braut dem Hochzeitsplaner den Schritt vorgetanzt hatte.</td><td>Der Bräutigam ahnte nicht, dass die Braut den Schritt dem Hochzeitsplaner vorgetanzt hatte.</td></tr>
<tr><td colspan="2">‘The groom did not suspect that the bride had danced the step for the wedding planner.’</td></tr>
<tr><td rowspan="3">42</td><td colspan="2">Beim heutigen Fußballtraining ging jeder Ball ins Netz. Offenbar war ein Hinweis darauf nötig, dass der Ball eigentlich abgewehrt werden sollte.
‘At today’s soccer training every ball scored. Apparently a hint was needed that the ball should actually be held off.’</td></tr>
<tr><td>Alle hofften, dass der Trainer dem Torwart den Hinweis zurufen würde.</td><td>Alle hofften, dass der Trainer dem Torwart den Hinweis zurufen würde.</td></tr>
<tr><td colspan="2">‘Everyone hoped that the coach would shout the hint to the goalkeeper.’</td></tr>
</table>

Comprehension questions

		Answer	**L1 accuracy**	**L2 accuracy**
27	Hat die Sekretärin den Kaufvertrag geemailt? ‘Did the secretary email the sales contract?’	No	45.2%	56.8%
29	Wurde der Betrag dem Waisenhaus gespendet? ‘Was the amount donated to the orphanage?’	No	100%	94.6%
31	Wollte das Paar einen Walzer vortanzen? ‘Did the couple want to dance a waltz?’	Yes	100%	100%

33	Hat der Leutnant den Befehl dem Kapitän zugerufen? 'Did the lieutenant shout the order to the captain?'	No	71%	62.2%
35	Wurde der Antrag dem Kollegen gefaxt? 'Was the proposal faxed to the colleague?'	Yes	96.8%	91.9%
36	Will der Rentner dem Finanzamt schreiben? 'Does the retiree want to write to the tax authorities?'	Yes	58.1%	70.3%
38	Hat die Mutter das Buch der Großmutter vorgelesen? 'Has the mother read the book to the grandmother?'	No	90.3%	86.5%
40	Durfte der Künstler einen Werbefilm vorspielen? 'Was the artist allowed to play a commercial?'	No	74.2%	64.9%

Practice and filler items

	Contextualizing sentences
	Sentence presented word-by-word
P1	Dienstags sind alle lange im Büro. Die Personalleiterin schlug vor, dass alle schneller arbeiten sollten. 'On Tuesdays everyone stays in the office until late. The staff executive suggested that everyone should work faster.'
	Es wurde spät und jeder wollte nach Hause gehen. 'It got late and everone wanted to go home.'
P2	Robert muss seinen Führerschein in der Verkehrskontrolle vorzeigen. Er ist viel zu schnell gefahren und hat auch noch getrunken. 'Robert has to present his license during a traffic control. He drove way too fast and has also drunken alcohol.'
	*Robert vorzeigt seinen Führerschein dem Polizisten nur zögerlich. 'Robert shows his license to the police officer only hesitantly.'
P3	Norbert arbeitete als verdeckter Ermittler. Auf einem gestohlenen PC entdeckte er geheime Bankdaten. 'Norbert works as an undercover agent. He found secret bank information on a stolen computer.'
	Es war unklar, ob der Ermittler den Computer der Polizei übergeben würde. 'It was unclear whether the agent would hand over the computer to the police.'
P4	Dennis darf den Kleiderschrank seiner Freundin vor dem Umzug zerlegen. Der Schrank ist sehr groß und sperrig. 'Dennis may disassemble the wardrobe of his girlfriend before the move. The wardrobe is very big and bulky.'
	Dennis legt den Kleiderschrank in den Transporter. 'Dennis puts the wardrobe into the transporter.'
43	Die kleine Charlotte hatte Leukämie. Verzweifelt suchten ihre Eltern eine Spezialistin auf. 'Little Charlotte has leukemia. Her parents desperately seek out a specialist.'
	Sie hofften sehr, dass die Ärztin dem Kind helfen konnte, wieder gesund zu werden. 'They truly hope that the doctor could help the child to get healthy again.'
44	Beim letzten Sturm war ein Baum auf das Haus von Familie Grünhof gefallen. Das Dach musste erneuert werden. 'During the last storm a tree had fallen onto the house of family Grünhof. The roof had to be renewed.'
	Das war sehr ärgerlich, weil die Familie das Haus erst letztes Jahr gekauft hatte.

	‘That was very annoying as the family had bought the house only last year’
45	Jeden Morgen gab es eine Besprechung in der Redaktion. Heute war die Stimmung schlecht. ‘Every morning there was a meeting in the editorial office. Today the mood was bad.’
	Der Nachrichtensprecher hatte keine Ahnung, warum die Zuschauer ihn langweilig fanden. ‘The news anchor had no idea why the viewers found him boring.’
46	Frau Schwarz hatte ein gutes Herz. Zufällig beobachtete sie, wie der Bettler sich eine Flasche Wodka kaufte. Mrs Schwarz had a good heart. She accidentally observed how the beggar bought himself a bottle of wodka.’
	Diesem Menschen würde sie ganz bestimmt kein Geld mehr geben. ‘She would surely not give money anymore to this man.’
47	Das Krankenhaus war voll belegt. Den Krankenschwestern war klar, dass es ein anstrengender Arbeitstag werden würde. ‘The hospital was fully occupied. It was clear to the nurses that it would be an exhausting work day.’
	Sie freuten sich schon sehr auf das Wochenende. ‘They already looked forward a lot to the weekend.’
48	Die große Fähre verließ den Hafen. Einer der Seemänner bemerkte, dass ein Kind ganz allein auf dem Deck stand. ‘The big ferry left the harbor. One of the sailors noticed a child standing on deck all alone.’
	Wahrscheinlich suchten die Eltern schon nach ihm. ‘The parents were probably already looking for him.’
49	Carolin muss den Mietwagen sofort bezahlen. Sie hat jedoch nicht genug Bargeld dabei. ‘Carolin had to pay immediately for the rental car. However she had not enough cash with her.’
	Carolin zahlt die Miete daher mit Kreditkarte. ‘Carolin therefore pays the rent with credit card.’
50	Tobias kann die Zahnarztpraxis einfach nicht betreten. Seine Mutter schiebt ihn durch die Tür. ‘Tobias simply cannot enter into the dental surgery. His mother pushes him through the door.’
	Tobias tritt seiner Mutter dabei auf die Füße. ‘In doing so Tobias steps onto his mother’s feet.’
51	Florian soll einen neuen Designersportwagen entwerfen. Er hat auch viele gute Ideen, besonders für die Farbe. ‘Florian is supposed to plan a new designer sports car. He has a lot of good ideas, especially for the color.’
	Florian wirft einen schlechten Entwurf in den Müll. ‘Florian throws a bad draft into the trash.’
52	Saskia will die Fotos von ihrem Ex-Freund zerschneiden. Er hat sie vor einer Woche verlassen. ‘Saskia wants to cut up the pictures of her former boyfriend. He left her a week ago.’
	Saskia schneidet alle Fotos in kleine Stücke. ‘Saskia cuts all pictures into small pieces.’
53	Andreas möchte sich gerne verlieben in diesem Frühjahr. In seiner Klasse sind viele nette Mädchen. ‘Andreas wants to fall in love this spring. In his class are many nice girls.’
	Andreas liebt eines der Mädchen über alles. ‘Andreas loves one of the girls more than anything.’

54	Nina wird sich in den engen Straßen der Altstadt bestimmt verfahren. Sie hört daher auf das Navigationsgerät. ‘Nina will surely lose her way in the narrow streets of the old town. She therefore listens to the satnav.’
	Nina fährt trotzdem in die falsche Gasse. ‘Nina drives into the wrong alley anyway.’
55	Meike will pünktlich zur ihrer ersten Fahrstunde losgehen. Dann ruft ihre Freundin an, um ihr Glück zu wünschen. ‘Meike wants to leave on time for her first driving lesson. Then her friend calls to wish her good luck.’
	*Meike losgeht mit einer Viertelstunde Verspätung. ‘Meike leaves with a delay of a quarter of an hour.’
56	Markus wird seinen Kumpel beim Fallschirmspringen nicht loslassen. Er hat viel zu große Höhenangst. ‘Markus does not want to let go of his buddy during skydiving. He is way too afraid of heights.’
	*Markus loslässt seinen Freund erst am Boden. ‘Markus lets go of his friend only on the ground.’
57	Jessica mag gerne bei der Theatergruppe mitspielen. Sie hat sich noch nie zum Vorsprechen getraut. ‘Jessica would like to play in the theater group. She has never dared to go to the audition.’
	*Jessica mitspielt dieses Jahr zum ersten Mal. ‘Jessica plays this year for the first time.’
58	Felix kann zu den Olympischen Spielen nach Russland mitkommen. Er hatte diese Saison immer gute Resultate. ‘Felix can tag along to Russia to the Olympic Games. He always had good results this season.’
	*Felix mitkommt auch zur Eröffnungsfeier am ersten Abend. ‘Felix also tags along to the opening ceremony on the first evening.’
59	Laura soll die Siegerehrung für das Herrenfinale vorbereiten. Sie ist für die Medaillen zuständig. ‘Laura shall prepare the victory ceremony for the men’s final. She is in charge of the medals’
	*Laura vorbereitet das Kissen, auf dem die Goldmedaille liegt. ‘Laura prepares the cushion on which the gold medal rests.’
60	Jasmin möchte den Sportlern beim Training zuschauen. Weil es regnet, findet das Training in der Halle statt. ‘Jasmin wants to watch the athletes during the training. As it is raining, the training takes place in the hall.’
	*Jasmin zuschaut den Sportlern dann in der Halle. ‘Jasmin then watches the athletes in the hall.’
61	Fabian darf die große Ausstellung am letzten Tag zuschließen. Sie war ein großer Erfolg und viele Besucher sind gekommen. ‘Fabian may lock up the big exhibition on the last day. It was a big success and many visitors came.’
	*Fabian zuschließt die Ausstellung mit ein wenig Trauer. ‘Fabian locks up the exihibit with some sorrow.’

Comprehension question

		answer
P4	Legt Dennis den Wickeltisch in den Transporter? ‘Does Dennis put the diaper changing table into the van?’	No

Study 3 – Norwegian particle verbs

Pilot study

Experimental items

	Short object	Mid-length object	Long object
01	Thomas river huset ned.	Thomas river det berømte huset ned.	Thomas river det berømte, gamle huset ned.
	'Thomas tears the house down.'	'… the famous house …'	'… the famous, old house …'
02	Lars trekker buksen ned.	Lars trekker den varme buksen ned.	Lars trekker den pene, varme buksen ned.
	'Lars pulls the pants down.'	'… the warm pants…'	'… the nice, warm pants…'
03	Kristian kaster søppelet bort.	Kristian kaster det stinkende søppelet bort.	Kristian kaster det ekle, stinkende søppelet bort.
	'Kristian throws the trash away.'	'…the stinking trash…'	'…the disgusting, stinking trash…'
04	Jan leier båten bort.	Jan leier den lille båten bort.	Jan leier den enkle, lille båten bort.
	'Jan rents the boat out.'	'… the small boat…'	'… the simple, small boat…'
05	Stian slår fjernsynet på.	Stian slår det store fjernsynet på.	Stian slår det store, brede fjernsynet på.
	'Stian turns the tv on.'	'… the big tv…'	'… the big, wide tv…'
06	Morten barberer skjegget av.	Morten barberer det lange skjegget av.	Morten barberer det lange, brune skjegget av.
	'Morten shaves the beard off.'	'…the long beard…'	'…the long, brown beard…'
07	Anders slipper hunden inn.	Anders slipper den blinde hunden inn.	Anders slipper den våte, blinde hunden inn.
	'Anders lets the dog in.'	'… the blind dog…'	'…the wet, blind dog…'
08	Ole taster passordet inn.	Ole taster det kompliserte passordet inn.	Ole taster det kompliserte, sikre passordet inn.
	'Ole types the password in.'	'…the complicated password…'	'…the complicated, secure password…'
09	Espen graver funnet ut.	Espen graver det arkeologiske funnet ut.	Espen graver det viktige, arkeologiske funnet ut.
	'Espen digs the find out.'	'… the archeological find…'	'…the important, archeological find…'
10	Øyvind fyller søknaden ut.	Øyvind fyller den deprimerende søknaden ut.	Øyvind fyller den siste deprimerende søknaden ut.
	'Øyvind fills the application out.'	'… the depressing application…'	'…the last, depressing application…'
11	Bjørn spiser kaken opp.	Bjørn spiser den gigantiske kaken opp.	Bjørn spiser den gigantiske lekre kaken opp.
	'Bjørn eats the cake up.'	'…the gigantic cake…'	'…the gigantic, tasty cake…'
12	Frode blåser ballongen opp.	Frode blåser den oransje ballongen opp.	Frode blåser den eneste oransje ballongen opp.
	'Frode blows the balloon	'…the orange balloon…'	'…the only orange

	up.'		balloon...'
13	Anne laster programmet ned.	Anne laster det aktualiserte programmet ned.	Anne laster det aktualiserte bedre programmet ned.
	'Anne downloads the program.'	'... the updated program...'	'... th updated, better program...'
14	Camilla skriver tanken ned.	Camilla skriver den alvorlige tanken ned.	Camilla skriver den intime alvorlige tanken ned.
	'Camilla writes the thought down.'	'... the serious thought...'	'...the intimate, serious thought...'
15	Kristin rydder leketøyet bort.	Kristin rydder det fantastiske leketøyet bort.	Kristin rydder det fantastiske nye leketøyet bort.
	'Kristin puts the toy away.'	'...the fantastic toy...'	'...the fantastic, new toy...'
16	Marte tauer bilen bort.	Marte tauer den ødelagte bilen bort.	Marte tauer den ødelagte billige bilen bort.
	'Marte tows the car away.'	'...the destroyed car...'	'...the destroyed, cheap car...'
17	Marianne setter ringen på.	Marianne setter den gylne ringen på.	Marianne setter den ekte gylne ringen på.
	'Marianne puts the ring on.'	'... the golden ring...'	'... the real golden ring...'
18	Silje tar jakken av.	Silje tar den blå jakken av.	Silje tar den tjukke blå jakken av.
	'Silje takes the jacket off.'	'...the blue jacket...'	'...the thick blue jacket...'
19	Linn puster luften inn.	Linn puster den arktiske luften inn.	Linn puster den kalde arktiske luften inn.
	'Linn breathes the air in.'	'... the Arctic air...'	'... the cold Arctic air...'
20	Ingrid pakker gaven inn.	Ingrid pakker den forferdelige gaven inn.	Ingrid pakker den forferdelige stygge gaven inn.
	'Ingrid wraps the gift up.'	'...the terrible gift...'	'...the terrible, ugly gift...'
21	Lene leser romanen ut.	Lene leser den spennende romanen ut.	Lene leser den spennende korte romanen ut.
	'Linn finishes the novel.'	'...the exciting novel...'	'...the exciting, short novel...'
22	Elin finner hemmeligheten ut.	Elin finner den fryktelige hemmeligheten ut.	Elin finner den fryktelige sjokkerende hemmeligheten ut.
	'Elin finds the secret out.'	'...the horrible secret...'	'...the horrible, shocking secret...'
23	Stine pusser rommet opp.	Stine pusser det mørke rommet opp.	Stine pusser det skitne mørke rommet opp.
	'Stine renovates the room.'	'...the dark room...'	'...the dirty, dark room...'
24	Hilde sier avtalen opp.	Hilde sier den overdrevne avtalen opp.	Hilde sier den overdrevne dyre avtalen opp.
	'Hilde withdraws from the contract.'	'...the exaggerated contract...'	'... the exaggerated expensive...'

Filler items

25	*Kristoffer drikker ølet ikke. 'Kristoffer does not drink the beer.'
26	*Rune åpner døren ikke. 'Rune does not open the door.'
27	*Tor vet svaret ikke. 'Tor does not know the answer.'
28	*Geir vekker barnet ikke. 'Geir does not wake the child.'
29	*Per vasker genseren ikke. 'Per does not wash the sweater.'
30	*Joakim vanner blomstene ikke. 'Joakim does not water the flowers.'
31	*Gunn planlegger reisen ikke. 'Gunn does not plan the journey.'
32	*Liv samler frimerker ikke. 'Liv does not collect stamps.'
33	*Katrine oversetter teksten ikke. 'Katrine does not translate the text.'
34	*Ragnhild overtar prosjektet ikke. 'Ragnhild does not take over the project.'
35	*Anette liker bildet ikke. 'Anette does not like the picture.'
36	*Ida lærer språket ikke. 'Ida does not learn the language.'

Experiment 3a – Acceptability rating task

Experimental items

	PO order	OP order
37	Thomas drikker opp den deilige rødvinen. 'Thomas drinks up the tasty red wine.'	Thomas drikker den deilige rødvinen opp. 'Thomas drinks the tasty red wine up.'
38	Anne henger opp det vakre bildet. 'Anne puts up the pretty picture.'	Anne henger det vakre bildet opp. 'Anne puts the pretty picture up.'
39	Camilla løfter opp det lille barnet. 'Camilla lifts up the small child.'	Camilla løfter det lille barnet opp. 'Camilla lifts the small child up.'
40	Kristian jager vekk de plagsomme myggene. 'Kristian chases away the annoying midges.'	Kristian jager de plagsomme myggene vekk. 'Kristian chases the annoying midges away.'
41	Stian slår av lyset. 'Stian turns off the light.'	Stian slår lyset av. 'Stian turns the light off.'
42	Marte slår opp informasjonen. 'Marte looks up the information.'	Marte slår informasjonen opp. 'Marte looks the information up.'
43	Marianne tar på de flotte skoene. 'Marianne puts on the fancy shoes.'	Marianne tar de flotte skoene på. 'Marianne puts the fancy shoes on.'
44	Anders trekker ut den betente tannen. 'Anders pulls out the inflamed tooth.'	Anders trekker den betente tannen ut. 'Anders pulls the inflamed tooth out.'

45	Ole fyller opp hyllen. 'Ole fills up the shelf.'	Ole fyller hyllen opp. 'Ole fills the shelf up.'
46	Linn tørker av tallerkenen. 'Linn dries off the plate.'	Linn tørker tallerkenen av. 'Linn dries the plate off.'
47	Ingrid slipper ut den temperamentsfulle hesten. 'Ingrid lets out the spirited horse.'	Ingrid slipper den temperamentsfulle hesten ut. 'Ingrid lets the spirited horse out.'
48	Rune dikter opp den dumme utflukten. 'Rune thinks up the stupid excuse.'	Rune dikter den dumme utflukten opp. 'Rune thinks the stupid excuse up.'
49	Tor setter inn pengene. 'Tor puts in the money.'	Tor setter pengene inn. 'Tor puts the money in.'
50	Liv viser fram passet. 'Liv shows the passport.'	Liv viser passet fram. 'Liv shows the passport.'
51	Katrine strekker ut tungen. 'Katrine sticks out the tongue.'	Katrine strekker tungen ut. 'Katrine sticks the tongue out.'
52	Jan pumper ut kjelleren. 'Jan pumps out the cellar.'	Jan pumper kjelleren ut. 'Jan pumps the cellar out.'

Filler items

26	*Lars åpner døren ikke. 'Lars does not open the door.'
29	*Morten vasker klærne aldri. 'Morten never washes the clothes.'
30	*Trond vanner blomstene sjelden. 'Trond rarely waters the flowers'.
25	*Monica drikker øl gjerne. 'Monica likes to drink beer.'
28	Kristin vekker ofte faren. 'Kristin often wakes the father.'
31	Silje planlegger stadig reiser. 'Silje constantly plans travels.'
53	I dag forstår endelig Ragnhild læreren. 'Today Ragnhild finally understands the teacher.'
54	Til jul besøker Geir alltid familien. 'On Christmas Geir always visits the family.'
55	Generelt overtar gjerne Gunn prosjekter. 'Generally, Gunn voluntarily takes over projects.'
56	Vanligvis oversetter Kristoffer sjelden bøker. 'Usually Kristoffer rarely translates books.'
57	*Terje lar seg skille fra Elisabeth. 'Terje gets a divorce from Elisabeth.'
58	*Knut slår seg med Kjell. 'Knut gives Kjell a beating.'
59	*Heidi laste boken. 'Heidi read the book.'
60	*Lene fåtte en gave. 'Lene got a present.'
61	*Julie bestemte seg om. 'Julie changed her mind.'
62	*Magnus tar i kurset del. 'Markus takes part in the class.'

APPENDIX A – Materials

Experiment 3b - SPR task

Experimental items

	PO_short PO_ long	OP_short OP_long
01	Thomas river ned huset tross protesten.	Thomas river huset ned tross protesten.
	Thomas river ned det berømte, gamle huset tross protesten.	Thomas river det berømte, gamle huset ned tross protesten.
	‘Thomas tears down the [famous, old] house despite the protest.’	‘Thomas tears the [famous, old] house down despite the protest.’
02	Lars trekker ned buksen på scenen.	Lars trekker buksen ned på scenen.
	Lars trekker ned den pene, varme buksen på scenen.	Lars trekker den pene, varme buksen ned på scenen.
	‘Lars pulls down the [nice, warm] pants on stage.’	‘Lars pulls the [nice, warm] pants down on stage.’
03	Kristian kaster bort søppelet i søppeldunken.	Kristian kaster søppelet bort i søppeldunken.
	Kristian kaster bort det ekle, stinkende søppelet i søppeldunken.	Kristian kaster det ekle, stinkende søppelet bort i søppeldunken.
	‘Kristian throws away the [disgusting, stinking] trash into the trashcan.’	‘Kristian throws the [disgusting, stinking] trash away into the trashcan.’
04	Jan leier bort båten til turistene.	Jan leier båten bort til turistene.
	Jan leier bort den enkle, lille båten til turistene.	Jan leier den enkle, lille båten bort til turistene.
	‘Jan rents out the [simple, small] boat to the tourists.’	‘Jan rents the [simple, small] boat out to the tourists.’
05	Stian slår på fjernsynet til fotballkampen.	Stian slår fjernsynet på til fotballkampen.
	Stian slår på det store, brede fjernsynet til fotballkampen.	Stian slår det store, brede fjernsynet på til fotballkampen.
	‘Stian turns on the [big, wide] tv for the soccer match.’	‘Stian turns the [big, wide] tv on for the soccer match.’
06	Morten barberer av skjegget etter bryllupet.	Morten barberer skjegget av etter bryllupet.
	Morten barberer av det lange, brune skjegget etter bryllupet.	Morten barberer det lange, brune skjegget av etter bryllupet.
	‘Morten shaves off the [long, brown] beard after the wedding.’	‘Morten shaves the [long, brown] beard off after the wedding.’
07	Anders slipper inn hunden av medlidenhet.	Anders slipper hunden inn av medlidenhet.
	Anders slipper inn den våte, blinde hunden av medlidenhet.	Anders slipper den våte, blinde hunden inn av medlidenhet.
	‘Andres lets in the [wet, blind] dog out of compassion.’	‘Andres lets the [wet, blind] dog in out of compassion.’
08	Ole taster inn passordet i mobiltelefonen.	Ole taster passordet inn i mobiltelefonen.
	Ole taster inn det kompliserte, sikre passordet i mobiltelefonen.	Ole taster det kompliserte, sikre passordet inn i mobiltelefonen.
	‘Ole types in the [complicated, secure] password into the cell phone.’	‘Ole types the [complicated, secure] password in into the cell phone.’
09	Espen graver ut funnet i ørkenen.	Espen graver funnet ut i ørkenen.
	Espen graver ut det viktige,	Espen graver det viktige, arkeologiske

	arkeologiske funnet i ørkenen.	funnet ut i ørkenen.
	'Espen digs out the [important, archeological] find in the desert.'	'Espen digs the [important, archeological] find out in the desert.'
10	Øyvind fyller ut søknaden i år.	Øyvind fyller søknaden ut i år.
	Øyvind fyller ut den siste deprimerende søknaden i år.	Øyvind fyller den siste deprimerende søknaden ut i år.
	'Øyvind fills out the [last, depressing] application this year.'	'Øyvind fills the [last, depressing] application out this year.'
11	Bjørn spiser opp kaken på bursdagsfesten.	Bjørn spiser kaken opp på bursdagsfesten.
	Bjørn spiser opp den gigantiske lekre kaken på bursdagsfesten.	Bjørn spiser den gigantiske lekre kaken opp på bursdagsfesten.
	'Bjørn eats up the [gigantic, tasty] cake at the birthday party.'	'Bjørn eats the [gigantic, tasty] cake up at the birthday party.'
12	Frode blåser opp ballongen for barnet.	Frode blåser ballongen opp for barnet.
	Frode blåser opp den eneste oransje ballongen for barnet.	Frode blåser den eneste oransje ballongen opp for barnet.
	'Frode blows up the [single orange] balloon for the child.'	'Frode blows the [single orange] balloon up for the child.'
13	Anne laster ned programmet fra internettet.	Anne laster programmet ned fra internettet.
	Anne laster ned det aktualiserte bedre programmet fra internettet.	Anne laster det aktualiserte bedre programmet ned fra internettet.
	'Anne downloads the [updated, better] program from the internet.'	
14	Camilla skriver ned tanken i dagboken.	Camilla skriver tanken ned i dagboken.
	Camilla skriver ned den intime alvorlige tanken i dagboken.	Camilla skriver den intime alvorlige tanken ned i dagboken.
	'Camilla writes down the [intimate, serious] thought in the diary.'	'Camilla writes the [intimate, serious] thought down in the diary.'
15	Kristin rydder bort leketøyet på loftet.	Kristin rydder leketøyet bort på loftet.
	Kristin rydder bort det fantastiske nye leketøyet på loftet.	Kristin rydder det fantastiske nye leketøyet bort på loftet.
	'Kristin puts away the [fantastic, new] toy in the attic.'	'Kristin puts the [fantastic, new] toy away in the attic.'
16	Marte tauer bort bilen fra ulykken.	Marte tauer bilen bort fra ulykken.
	Marte tauer bort den ødelagte billige bilen fra ulykken.	Marte tauer den ødelagte billige bilen bort fra ulykken.
	'Marte tows away the [destroyed, cheap] car from the accident.'	'Marte tows the [destroyed, cheap] car away from the accident.'
17	Marianne setter på ringen i kirken.	Marianne setter ringen på i kirken.
	Marianne setter på den ekte gylne ringen i kirken.	Marianne setter den ekte gylne ringen på i kirken.
	'Marianne puts on the [real golden] ring in church.'	'Marianne puts the [real golden] ring on in church.'
18	Silje tar av jakken på hytten.	Silje tar jakken av på hytten.
	Silje tar av den tjukke blå jakken på hytten.	Silje tar den tjukke blå jakken av på hytten.
	'Silje takes off the [thick, blue] jacket in the cabin.'	'Silje takes the [thick, blue] jacket off in the cabin.'
19	Linn puster inn luften på Svalbard.	Linn puster luften inn på Svalbard.
	Linn puster inn den kalde arktiske luften på Svalbard.	Linn puster den kalde arktiske luften inn på Svalbard.
	'Linn breathes in the [cold Arctic] air	'Linn breathes the [cold Arctic] air in on

	on Svalbard.'	Svalbard.'
20	Ingrid pakker inn gaven i papir.	Ingrid pakker gaven inn i papir.
	Ingrid pakker inn den forferdelige stygge gaven i papir.	Ingrid pakker den forferdelige stygge gaven inn i papir.
	'Ingrid wraps up the [terrible, ugly] gift in paper.'	'Ingrid wraps the [terrible, ugly] gift up in paper.'
21	Lene leser ut romanen før eksamen.	Lene leser romanen ut før eksamen.
	Lene leser ut den spennende korte romanen før eksamen.	Lene leser den spennende korte romanen ut før eksamen.
	'Linn finishes the [exciting, short] novel before the exam.'	
22	Elin finner ut hemmeligheten på kirkegården.	Elin finner hemmeligheten ut på kirkegården.
	Elin finner ut den fryktelige sjokkerende hemmeligheten på kirkegården.	Elin finner den fryktelige sjokkerende hemmeligheten ut på kirkegården.
	'Elin finds out the [horrible, shocking] secret on the graveyard.'	'Elin finds the [horrible, shocking] secret out on the graveyard.'
23	Stine pusser opp rommet etter inflyttingen.	Stine pusser rommet opp etter inflyttingen.
	Stine pusser opp det skitne mørke rommet etter inflyttingen.	Stine pusser det skitne mørke rommet opp etter inflyttingen.
	'Stine renovates the [dirty, dark] room after moving in.'	
24	Hilde sier opp avtalen til desember.	Hilde sier avtalen opp til desember.
	Hilde sier opp den overdrevne dyre avtalen til desember.	Hilde sier den overdrevne dyre avtalen opp til desember.
	'Hilde withdraws from the [exaggerated, expensive] contract by december.'	

Comprehension questions

		Answer	L1 accuracy	L2 accuracy
03	Kaster Kristian søppelet i søppeldunken? 'Does Kristian throw the trash into the trashcan?'	Yes	90.63%	93.75%
05	Er det computeren Stian slår på? 'Is it the computer that Stian turns on?'	No	100%	93.75%
08	Er passordet til Oles telefonen? 'Is it the password for Ole's phone?'	Yes	93.75%	78.13%
14	Skriver Camilla i avisen? 'Does Camilla write in the newspaper?'	No	100%	100%
15	Er leketøyet på loftet? 'Is the toy in the attic?'	Yes	75%	71.88%
16	Er det motorsyklen Marte tauer bort? 'Is it the motorcycle that Marte tows away?'	No	96.88%	96.88%
18	Tar Silje jakken av på hytten? 'Does Silje take off the jacket in the cabin?'	Yes	96.88%	100%
19	Er Linn på Mallorca? 'Is Linn in Mallorca?'	No	100%	100%

Study 4 –German particle verbs

Experiment 4a - Acceptability rating task

Experimental items

	Split	**No split**
01	Die Bombe reißt dem Attentäter das Bein ab.	*Die Bombe abreißt dem Attentäter das Bein.
	‘The bomb tears off the leg of the assassin.’	
02	Der Chemiker nimmt die Schutzbrille ab.	*Der Chemiker abnimmt die Schutzbrille.
	‘The chemist takes off the safety glasses.’	
03	Der Obdachlose wirft Pfandflaschen weg.	*Der Obdachlose wegwirft Pfandflaschen.
	‘The homeless person throws away returnable bottles.’	
04	Der Student leiht das Kursbuch aus.	*Der Student ausleiht das Kursbuch.
	‘The student borrows the textbook.’	
05	Der Chef stellt eine neue Sekretärin an.	*Der Chef anstellt eine neue Sekretärin.
	‘The boss hires a new secretary.’	
06	Der Frisör rasiert dem Soldaten die Haare ab.	*Der Frisör abrasiert dem Soldaten die Haare.
	‘The hairdresser cuts off the hairs of the soldier.’	
07	Der Türsteher lässt keine weiteren Gäste ein.	*Der Türsteher einlässt keine weiteren Gäste.
	‘The bouncer does not let in any further guests.’	
08	Der Bankangestellte gibt den Code für den Tresor ein.	*Der Bankangestellte eingibt den Code für den Tresor.
	‘The bank employee enters the code for the safe.’	
09	Die Zeitung gräbt wichtige Informationen aus.	*Die Zeitung ausgräbt wichtige Informationen.
	‘The newspaper digs up important information.’	
10	Der Spieler füllt zehn Lottoscheine aus.	*Der Spieler ausfüllt zehn Lottoscheine.
	‘The gambler fills out ten lottery tickets.’	
11	Das Kind isst seinen Brei auf.	*Das Kind aufisst seinen Brei.
	‘The child eats up his porridge.’	
12	Der Clown bläst die Backen auf.	*Der Clown aufbläst die Backen.
	‘The clown blows up his cheeks.’	
13	Das Mädchen fasst vorsichtig das Pferd an.	*Das Mädchen anfasst vorsichtig das Pferd.
	‘The girl carefully touches the horse.’	
14	Der Falschparker fährt sein Auto weg.	*Der Falschparker wegfährt sein Auto.
	‘The parking offender drives his car away.’	
15	Die Familie räumt den Dachboden auf.	*Die Familie aufräumt den Dachboden.
	‘The family tidies up the attic.’	
16	Die Polizei schleppt das Auto ab.	*Die Polizei abschleppt das Auto.
	‘The police tows away the car.’	
17	Der König setzt die Krone auf.	*Der König aufsetzt die Krone.
	‘The king puts on the crown.’	
18	Der Stripper zieht seine Kleidung aus.	*Der Stripper auszieht seine Kleidung.

	'The stripper takes off his clothes.'	
19	Der Feuerwehrmann atmet giftige Dämpfe ein.	*Der Feuerwehrmann einatmet giftige Dämpfe.
	'The firefighter breates in toxic fumes.'	
20	Die Verkäuferin packt die Waren ein.	*Die Verkäuferin einpackt die Waren.
	'The shop assistant wraps up the goods.'	
21	Der Millionär gibt sein ganzes Vermögen weg.	*Der Millionär weggibt sein ganzes Vermögen.
	'The millionaire gives away his entire fortune.'	
22	Das Gericht hört den Zeugen an.	*Das Gericht anhört den Zeugen.
	'The court listens to the witness.'	
23	Der Bauer baut Biokartoffeln an.	*Der Bauer anbaut Biokartoffeln.
	'The farmer cultivates organic potatoes.'	
24	Der Schiedsrichter sagt den Olympischen Eid auf.	*Der Schiedsrichter aufsagt den Olympischen Eid.
	'The referee recites the Olympic oath.'	

Experiment 4b – SPR task

Experimental items

	Contextualizing sentences	
	Auxiliary_no split	**Embedded_no split**
	Auxiliary_split	**Embedded_ split**
01	Julia reißt ungern das Gartenhaus ihrer Eltern ab. Es ist alt und verfallen.	
	'Julia unwillingly demolishes the summer house of her parents. It is old and deteriorated.'	
	Julia wird das alte Gartenhaus abreißen nächstes Wochenende.	Die Nachbarn bedauern, dass Julia das alte Gartenhaus abreißt nächstes Wochenende.
	*Julia wird das alte Gartenhaus reißen ab nächstes Wochenende.	*Die Nachbarn bedauern, dass Julia das alte Gartenhaus reißt ab nächstes Wochenende.
	'Julia will demolish the old summer house next weekend.'	'The neighbors regret that Julia will demolish the old summer house next weekend.'
02	Sebastian nimmt jeden Monat fünf Kilo ab. Insgesamt sind es schon 25 Kilo, aber er ist immer noch sehr schwer. 'Sebastian loses five kilograms each month. Altogether it is already 25 kilograms, but he is still very heavy.'	
	Sebastian soll nochmal fünfzehn Kilo abnehmen diesen Sommer.	Die Ärztin verordnet, dass Sebastian nochmal fünfzehn Kilo abnimmt diesen Sommer.
	*Sebastian soll nochmal fünfzehn Kilo nehmen ab diesen Sommer.	*Die Ärztin verordnet, dass Sebastian nochmal fünfzehn Kilo nimmt ab diesen Sommer.
	'Sebastian shall lose another fifteen kilograms this summer.'	'The doctor prescribes, that Sebastian loses another fifteen kilograms this summer.'
03	Sarah wirft niemals etwas weg. Ihre ganze Wohnung ist voll mit Müll. 'Sarah never throws anything away. Her entire apartment is full with trash.'	
	Sarah möchte alle Zeitungen wegwerfen nach ihrer Therapie.	Ihr Psychologe empfiehlt, dass Sarah alle Zeitungen wegwirft nach ihrer Therapie.
	*Sarah möchte alle Zeitungen werfen	*Ihr Psychologe empfiehlt, dass Sarah alle

	weg nach ihrer Therapie.	Zeitungen wirft weg nach ihrer Therapie.
	'Sarah wants to throw away all newspapers after her therapy.'	'Her psychologist recommends that Sarah throws away all newspapers after her therapy.'
04	Alexander leiht gerne Bücher aus in der Bibliothek. Er vergisst aber oft, sie pünktlich zurückzubringen. 'Alexander likes to borrow books from the library. But he often forgets to return them on time.'	
	Alexander darf keine Bücher mehr ausleihen in Zukunft.	Die Bibliothekarin entscheidet, dass Alexander keine Bücher mehr ausleiht in Zukunft.
	*Alexander darf keine Bücher mehr leihen aus in Zukunft.	*Die Bibliothekarin entscheidet, dass Alexander keine Bücher mehr leiht aus in Zukunft.
	'Alexander is not allowed to borrow any more books in the future.'	'The librarian decides that Alexander does not borrow any more books in the future.'
05	Melanie stellt die Heizung selten an. Die Heizkosten sind für sie zu hoch. 'Melanie seldom turns on the heating. The heating costs are just too high for her.'	
	Melanie muss die Heizung aber anstellen im Winter.	Der Vermieter verlangt, dass Melanie die Heizung aber anstellen im Winter.
	*Melanie muss die Heizung aber stellen an im Winter.	*Der Vermieter verlangt, dass Melanie die Heizung aber stellen an im Winter.
	'Melanie has to turn on the heating in winter though.'	'The landlord demands that Melanie turns on the the heating in winter though.'
06	Jan rasiert seine Brusthaare normalerweise nicht ab. Seine neue Freundin findet das nicht so toll. 'Jan normally does not shave off his chest hair. His new girlfriend does not think that great.'	
	Jan wird seine Brusthaare abrasieren von morgen an.	Die Freundin befiehlt, dass Jan seine Brusthaare abrasiert von morgen an.
	*Jan wird seine Brusthaare rasieren ab von morgen an.	*Die Freundin befiehlt, dass Jan seine Brusthaare rasiert ab von morgen an.
	'Jan will shave off his chest hair from tomorrow on.'	'The girlfriend commands that Yan shaves off his chest hair from tomorrow on.'
07	Jennifer lässt normalerweise keine Missionare ein in ihre Wohnung. Es regnet heute und ist kalt. 'Jennifer normally does not let in any missionaries into her apartment. Today it is raining and it is cold.'	
	Jennifer mag die Missionare einlassen in die Wohnung.	Der Pfarrer veranlasst, dass Jennifer die Missionare einlässt in die Wohnung.
	*Jennifer mag die Missionare lassen ein in die Wohnung.	*Der Pfarrer veranlasst, dass Jennifer die Missionare lässt ein in die Wohnung.
	'Jennifer wants to let in the missionaries into the apartment.'	'The priest determines that Jennifer lets in the missionaries into the apartment.'
08	Martin gibt seine Daten immer schnell in den Computer ein. Heute hat er wenig Zeit. 'Martin always enters his data quickly into the computer. Today he has little time.'	
	Martin wird die Daten schneller eingeben heute Morgen.	Sein Betreuer will, dass Martin die Daten schneller eingibt heute Morgen.
	*Martin wird die Daten schneller geben ein heute Morgen.	*Sein Betreuer will, dass Martin die Daten schneller gibt ein heute Morgen.
	'Martin will enter the data more quickly this morning.'	'His supervisor wants that Martin enters the data more quickly this morning.'
09	Katharina gräbt bei Exkursionen immer etwas aus. Auch in diesem Semester fährt	

	sie auf eine Grabung. 'Katharaina always digs out something during excursions. She goes on a dig also this semester.'	
	Katharina kann vielleicht eine Mumie ausgraben in Ägypten.	Ihr Professor scherzt, dass Katharina vielleicht eine Mumie ausgräbt in Ägypten.
	*Katharina kann vielleicht eine Mumie graben aus in Ägypten.	*Ihr Professor scherzt, dass Katharina vielleicht eine Mumie gräbt aus in Ägypten.
	'Katharina can possibly dig out a mummy in Egypt.'	'Her professor jokes that Katharina can possibly dig out a mummy in Egypt.'
10	Daniel füllt viele Bewerbungen für Stipendien aus. Bisher war keine erfolgreich. 'Daniel fills out many applications for scholarships. So far none was successful.'	
	Daniel darf noch zehn Bewerbungen ausfüllen bis Dezember.	Sein Doktorvater verlangt, dass Daniel noch zehn Bewerbungen ausfüllt bis Dezember.
	*Daniel darf noch zehn Bewerbungen füllen aus bis Dezember.	*Sein Doktorvater verlangt, dass Daniel noch zehn Bewerbungen füllt aus bis Dezember.
	'Daniel may fill out an additional ten applications until December.'	'His Ph.D. supervisor demands that Daniel fills out an additional ten applications until December.'
11	Anne isst beim Mittagessen brav ihren Salat auf. Sie ist aber immer noch hungrig. 'Anne well-behavedly eats up her salad during lunch. But she is still hungry.'	
	Anne möchte noch ein Schnitzel aufessen im Anschluss.	Ihr Freund rät, dass Anne noch ein Schnitzel aufisst im Anschluss.
	*Anne möchte noch ein Schnitzel essen auf im Anschluss.	*Ihr Freund rät, dass Anne noch ein Schnitzel isst auf im Anschluss.
	'Anne wants to eats up a schnitzel afterwards.'	'Her boyfriends suggests that Anne eats up a schnitzel afterwards.'
12	Philipp bläst nur sehr ungern die Luftballons für seine Tochter auf. Heute hat Philipps Tochter Geburtstag. 'Philipp only reluctant blows up the balloons for his daughter. Today is the birthday of Philipp's daughter.	
	Philipp muss die Luftballons aufblasen vor der Feier.	Seine Tochter möchte, dass Philipp die Luftballons aufbläst vor der Feier.
	*Philipp muss die Luftballons blasen auf vor der Feier.	*Seine Tochter möchte, dass Philipp die Luftballons bläst auf vor der Feier.
	'Philipp has to blow up the balloons before the party.'	'His daughter wants that Philipp blows up the balloons before the party.'
13	Christina fasst problemlos Frösche an im Labor. Sie ist letzten Freitag Mutter geworden. 'Christina touches frogs in the lab without problems. She had a baby last Friday.'	
	Christina mag die Windeln nicht anfassen seit Freitag.	Die Krankenschwester versteht, dass Christina die Windeln nicht anfasst seit Freitag.
	*Christina mag die Windeln nicht fassen an seit Freitag.	*Die Krankenschwester versteht, dass Christina die Windeln nicht fasst an bisher seit Freitag.
	'Christina does not want to touch the diapers since Friday.'	'The nurse understands that Christina does not touch the diapers since Friday.'
14	Patrick fährt an Ostern immer weg. Am liebsten fährt er an einen warmen Ort am Meer.	
	Patrick always goes away on Easter. He likes best to go to a warm place on the sea.'	
	Patrick will dieses Jahr wegfahren in	Das Reisebüro offeriert, dass Patrick

	die Emirate.	dieses Jahr wegfährt in die Emirate.
	*Patrick will dieses Jahr bald fahren weg in die Emirate.	*Das Reisebüro offeriert, dass Patrick dieses Jahr fährt weg in die Emirate.
	‘Patrick wants to go away to the Emirates this year.’	‘The travel agency offers that Patrick goes away to the Emirates this year.’
15	Sandra räumt ständig das Spielzeug der Kindergartengruppe weg. Die Kinder haben keine Lust darauf. ‘Sandra constantly tidies up the toys of the nursery group. The children do not feel like it.’	
	Sandra will das Spielzeug nicht mehr wegräumen heute.	Ihre Kollegin beschließt, dass Sandra das Spielzeug nicht mehr wegräumt heute.
	*Sandra will das Spielzeug nicht mehr räumen weg heute.	*Ihre Kollegin beschließt, dass Sandra das Spielzeug nicht mehr räumt weg heute.
	‘Sandra does not want to tidy up the toys anymore today.’	‘Her colleague decides that Sandra does not tidy up the toys anymore today.
16	David schleppt kaputte Autos ab. Heute hat er viel zu tun. ‘David tows broken cars. He has a lot to do today.’	
	David kann den Honda nicht abschleppen vor Schichtende.	Sein Chef sagt, dass David den Honda nicht abschleppt vor Schichtende.
	*David kann den Honda nicht schleppen ab vor Schichtende.	*Sein Chef sagt, dass David den Honda nicht schleppt ab vor Schichtende.
	‘David cannot tow the Honda before the end of the shift.’	‘His boss says that David does not tow the Honda before the end of the shift.’
17	Sabrina setzt meistens keine Hüte auf. Heute werden allerdings die Familienfotos gemacht. ‘Sabrina mostly does not put on hats. Today however the family portraits are taken.’	
	Sabrina möchte einen Sonnenhut aufsetzen für das Foto.	Ihr Ehemann drängt, dass Sabrina einen Sonnenhut aufsetzt für das Foto.
	*Sabrina möchte einen Sonnenhut setzen auf für das Foto.	*Ihr Ehemann drängt, dass Sabrina einen Sonnenhut setzt auf für das Foto.
	‘Sabrina wants to put on a sun hat for the picture.’	‘Her husband pushes that Sabrina puts on a sun hat for the picture.’
18	Matthias zieht nie seine Schuhe aus beim Betreten einer Wohnung. Heute sind seine Schuhe sehr dreckig. ‘Matthias never takes off his shoes when entering into an apartment. Today his shoes are very dirty.’	
	Matthias soll die Schuhe sofort ausziehen im Flur.	Seine Schwiegermutter fordert, dass Matthias die Schuhe sofort auszieht im Flur.
	*Matthias soll die Schuhe sofort ziehen aus im Flur.	*Seine Schwiegermutter fordert, dass Matthias die Schuhe sofort zieht aus im Flur.
	‘Matthias shall take off the shoes immediately in the hall.’	‘His mother-in-law demands that Matthias takes off the shoes immediately in the hall.’
19	Nadine atmet auf ihrem Chinaurlaub viele Abgase ein. Nach ihrer Rückkehr nach Deutschland hat sie Atembeschwerden. ‘Nadine breathes in plenty of emissions during her holiday in China. After her return to Germany she has breathing problems.’	
	Nadine soll viel frische Waldluft einatmen auf Spaziergängen.	Der Arzt verschreibt, dass Nadine viel frische Waldluft einatmet auf Spaziergängen.
	*Nadine soll viel frische Waldluft atmen ein auf Spaziergängen.	*Der Arzt verschreibt, dass Nadine viel frische Waldluft atmet ein auf

		Spaziergängen.
	‘Nadine shall breathe in plenty of fresh wood air during walks.’	‘The doctor prescribes that Nadine breathes in plenty of fresh wood air during walks.’
20	Michael packt die Weihnachtsgeschenke immer sofort ein. Er vergisst dann den Inhalt. ‘Michael always wraps the Chritsmas present immediately. He then forgets the content’	
	Michael muss die Geschenke nochmal einpacken vorm Fest.	Seine Großmutter bittet, dass Michael die Geschenke nochmal einpackt vorm Fest.
	*Michael muss die Geschenke nochmal packen ein vorm Fest.	*Seine Großmutter bittet, dass Michael die Geschenke nochmal packt ein vorm Fest.
	‘Michael has to wrap the presents again before the celebration.’	‘His grandmother asks that Michael wraps the presents again before the celebration.’
21	Annika gibt die kleinen Hunde an Freunde weg. Den kleinen Braunen findet sie besonders süß. ‘Annika gives away the little dogs to friends. She thinks the little brown one is especially cute.’	
	Annika mag den Hund nicht weggeben bis Juli**.**	Ihre Mutter erlaubt, dass Annika den Hund nicht weggibt bis Juli.
	*Annika mag den Hund nicht geben weg bis Juli.	*Ihre Mutter erlaubt, dass Annika den Hund nicht gibt weg bis Juli.
	‘Annika does not want to give away the dog until July.’	‘Her mother allows that Annika does not give away the dog until July.’
22	Thomas hörte ein klassisches Konzert an letztes Jahr. Normalerweise hört er nur Rockmusik. ‘Thomas listened to a classical concert last year. Usually he only listenes to rock music.’	
	Thomas will diesmal eine Oper anhören im Theater.	Der Musiklehrer bestimmt, dass Thomas diesmal eine Oper anhört im Theater.
	*Thomas will diesmal eine Oper hören an im Theater.	*Der Musiklehrer bestimmt, dass Thomas diesmal eine Oper hört an im Theater.
	‘Thomas wants to listen to an opera in the theater this time.’	‘The music theater determines that Thomas listens to an opera in the theater this time.’
23	Franziska baut Pflanzen an in ihrem Keller. Eines Tages steht die Polizei vor ihrer Tür. ‘Franziska cultivates plants in her cellar. One day the police is at her door.’	
	Franziska darf kein Cannabis mehr anbauen im Keller.	Der Richter verfügt, dass Franziska kein Cannabis mehr anbaut im Keller.
	*Franziska darf kein Cannabis mehr bauen an im Keller.	*Der Richter verfügt, dass Franziska kein Cannabis mehr baut an im Keller.
	‘Franziska may no longer cultivate cannabis in the cellar.’	‘The judge rules that Franzika no longer cultivates cannabis in the cellar.’
24	Christoph sagt das Gedicht fehlerfrei auf. Er möchte ein neues Gedicht lernen. ‘Christoph recites the poem without mistakes. He wants to learn a new poem.’	
	Christoph kann die Glocke aufsagen für die Schulfeier.	Der Schuldirektor prahlt, dass Christoph die Glocke aufsagt für die Schulfeier.
	*Christoph kann die Glocke sagen auf für die Schulfeier.	*Der Schuldirektor prahlt, dass Christoph die Glocke sagt auf für die Schulfeier.
	‘Christoph can recite ‘Die Glocke’ for the school festivity.’	‘The principal boasts that Christoph recites ‘Die Glocke’ for the school festivity.’

Comprehension questions

		Answer	L1 accuracy	L2 accuracy
01	Demoliert Julia das alte Gartenhaus am Wochenende? 'Does Julia demolish the old summer house on the weekend?'	Yes	41.9%	56.8%
02	Soll Sebastian fünfzehn Kilo zunehmen? 'Shall Sebastian gain fifteen kilograms?'	No	90.3%	83.8%
03	Entsorgt Sarah alle Zeitungen nach der Therapie? 'Does Sarah dispose of all the newspapers after the therapy?'	Yes	63.3%	63.9%
04	Darf Alexander noch Bücher ausleihen? 'May Alexander still borrow books?'	No	100%	88.9%
05	Muss Melanie im Winter heizen? 'Does Melanie have to heat?'	Yes	93.5%	91.9%
06	Entfernt Jan seine Brusthaare morgen? 'Does jan remove his chest hair tomorrow?'	Yes	93.5%	85.7%
07	Lässt Jennifer die Missionare in die Wohnung? 'Does Jennifer let in missionaries into the apartment?'	Yes	93.5%	75.7%
08	Gibt Martin die Daten heute schnell ein? 'Does Martin enter the date quickly today?'	Yes	74.2%	78.4%
09	Gräbt Katharina eine Mumie ein in Ägypten? 'Does Katharina dig up a mummy in Egypt?'	No	96.8%	75.7%
10	Füllt Daniel Bewerbungen aus? 'Does Daniel fill out applications?'	Yes	64.5%	86.5%
11	Möchte Anne noch ein Schnitzel aufessen? 'Does Anne want to eat up a schnitzel?'	Yes	61.3%	78.4%
12	Füllt Philipp Ballons mit Luft? 'Does Philipp fill balloons with air?'	Yes	83.9%	75.7%
17	Setzt Sabrina den Sonnenhut ab für das Foto? 'Does Sabrina take off the sun hat for the picture?'	No	96.7%	70.3%
18	Zieht Matthias die Schuhe an im Flur? 'Does Matthias put on the shoes in the hall?'	No	90.3%	75.7%
19	Soll Nadine frische Waldluft ausatmen? 'Shall Nadine breathe out fresh wood air?'	No	83.9%	64.9%
20	Packt Michael die Geschenke aus vor dem Fest? 'Does Michael unpack the gifts before the celebration?'	No	71%	51.4%

APPENDIX B – Additional data

Study 1 – Norwegian object topicalizations

Pilot study

Item	$M_{animate}$	$SD_{animate}$	$M_{inanimate}$	$SD_{inanimate}$	Δ anim. – inan.
2	2.14	1.38	2.26	1.43	-0.12
4	1.04	0.19	1.10	0.41	-0.07
5	2.03	1.27	1.22	0.64	0.81
6	1.22	0.58	2.97	1.38	-1.74
8	3.41	1.50	2.04	1.13	1.38
9	1.81	1.14	1.93	1.33	-0.12
11	1.86	1.19	2.56	1.48	-0.69
12	2.44	1.09	3.62	1.45	-1.18
13	1.24	0.69	3.33	1.75	-2.09
14	2.59	1.37	2.48	1.53	0.11
15	2.14	1.16	3.37	1.28	-1.23
16	2.41	1.53	1.21	0.49	1.20
17	3.86	1.43	3.33	1.52	0.53
18	1.15	0.46	1.72	1.16	-0.58
19	4.38	0.86	1.67	0.83	2.71
20	1.89	1.05	4.14	1.09	-2.25
21	1.55	1.09	3.04	1.43	-1.49
22	3.26	1.16	2.97	1.24	0.29
23	1.86	1.22	1.67	1.11	0.20
24	1.69	0.97	4.93	0.27	-3.24
25	1.15	0.36	3.03	1.43	-1.89
26	1.52	0.94	2.69	1.31	-1.17
27	4.33	1.00	4.59	0.87	-0.25
28	3.90	1.45	3.33	1.21	0.56
29	1.41	0.89	3.03	1.57	-1.63
31	1.93	1.14	4.59	0.87	-2.66
32	1.72	1.03	3.04	1.29	-1.31
33	4.38	1.24	3.81	1.14	0.56
34	1.21	0.41	2.41	1.31	-1.20
35	1.31	0.54	4.19	1.04	-2.87
36	1.70	0.99	1.45	0.91	0.26
37	1.66	1.32	3.26	1.53	-1.60
38	1.52	1.02	1.89	1.25	-0.37
39	1.15	0.46	2.38	1.57	-1.23
40	1.26	0.66	2.38	1.40	-1.12
41	1.83	1.26	1.11	0.32	0.72

42	1.19	0.48	1.34	0.55	-0.16

Table B1.1 Average plausibility ratings for all 37 items used in the pilot study of Study 1, excluded items shaded

Experiment 1a - Agent identification task

Item	Condition	Animacy status	L1 Accuracy	L2 Accuracy
08	SVO	Inanimate subject	93.75%	100%
	OVS		50%	62.5%
19	SVO	Inanimate subject	93.75%	100%
	OVS		50%	75%
13	SVO	Animate subject	100%	100%
	OVS		75%	56.25%
35	SVO	Animate subject	100%	100%
	OVS		81.25%	68.75%
24	SVO	Animate subject	100%	100%
	OVS		81.25%	81.25%
31	SVO	Animate subject	100%	100%
	OVS		87.5%	87.5%
50	SVO	Equal animacy	100%	100%
	OVS		68.75%	31.25%
51	SVO	Equal animacy	87.5%	100%
	OVS		68.75%	62.5%
52	SVO	Equal animacy	100%	100%
	OVS		50%	56.25%
53	SVO	Equal animacy	100%	100%
	OVS		56.25%	62.5%
54	SVO	Equal animacy	93.75%	93.75%
	OVS		56.25%	68.75%
55	SVO	Equal animacy	81.25%	100%
	OVS		56.25%	68.75%
01	OVS	Semantic disambiguation	12.5%	68.75%
56	OVS	Semantic disambiguation	25%	50%
57	SVO	Semantic disambiguation	96.87%	100%
58	SVO	Semantic disambiguation	100%	100%
59	SVO	Cleft sentence	96.87%	100%
60	OVS	Cleft sentence	87.5%	87.5%
61	SVO	Cleft sentence	100%	96.87%
62	OVS	Cleft sentence	90.62%	93.75%
63	SVO	Cleft sentence	96.87%	100%
64	OSV	Cleft sentence	90.62%	87.5%
65	SVO	Cleft sentence	96.87%	100%
66	OVS	Cleft sentence	87.5%	84.37%
67	XVSO	Inversion	100%	96.87%
68	XVSO	Inversion	96.87%	100%
69	XVSO	Inversion	100%	96.87%
70	XVSO	Inversion	96.87%	100%

Table B1.2 By-item accuracy scores (no participants excluded)

Experiment 3b – SPR task

Segment	Extreme value cutoff	Removed data	Percentage

		points	
1 (NP1)	upper: 5000ms	50	3.46%
2 (relative pronoun)	upper: 3000ms	50	3.46%
3 (adverb)	upper: 3000ms	41	2.84%
4 (verb)	upper: 3000ms	38	2.63%
5 (adjective)	upper: 5000ms	42	2.91%
6 (auxiliary)	upper: 2500ms	51	3.53%
7 (main verb/NP2)	upper: 3500ms	50	3.46%
8 (NP2/main verb)	upper: 5000ms	46	3.19%
9 (preposition)	upper: 2500ms	48	3.32%
10 (NP)	upper: 12000ms	46	3.19%

Table B1.3 Data cleaning procedure of Experiment 1b

Segment	**F1-ANOVA (1,60)**	**F2-ANOVA (1,23)**
1 (NP1)	ME of Group F=15.89, p<0.001	ME of Group F=143.76, p<0.001
2 (relative pronoun)	Group *F=3.95, p=0.051*	ME of Group F=58.93, p<0.001
3 (adverb)	ME of Group F=8.87, p=0.004	ME of Group F=89.61, p<0.001
4 (verb)	ME of Group F =5.9, p=0.018	ME of Group F=45.46, p<0.001
5 (adjective)	ME of Group F=12.25, p<0.001	ME of Group F=201.62, p<0.001 Order *F=3.82, P=0.063*
6 (auxiliary)	ME of Group F=9.64, p=0.003	ME of Group F=151.1, p<0.001
7 (main verb/NP2)	ME of Group F=14.28, p<0.001 ME of Order F=44.84, p<0.001	ME of Group F=221.89, p<0.001 ME of Order F=13.68, p=0.0012
8 (NP2/main verb)	ME of Group F=17.36, p<0.001 ME of Animacy F=6.1 p=0.016 Animacy x Order F=7.76, p=0.007 Group x Order x Animacy *F=3.01, p=0.088*	ME of Group F=149.64, p<0.001 ME of Animacy F=5.82, p=0.024 Animacy x Order F=4.89, p=0.037
9 (preposition)	ME of Group F=7.93, p=0.007	ME of Group F=75.4, p<0.001 Group x Animacy *F=4.63, p=0.08*
10 (NP)	ME of Group *F=3.86, p=0.054* ME of Order F=23.54, p=<0.001 Group x Order *F=3.3, p=0.074*	ME of Group F=105.84, p<0.001 ME of Order F=6.4, p=0.02

Table B1.4 Between-groups ANOVAs per segment for log-transformed RTS (only significant or marginally significant values)

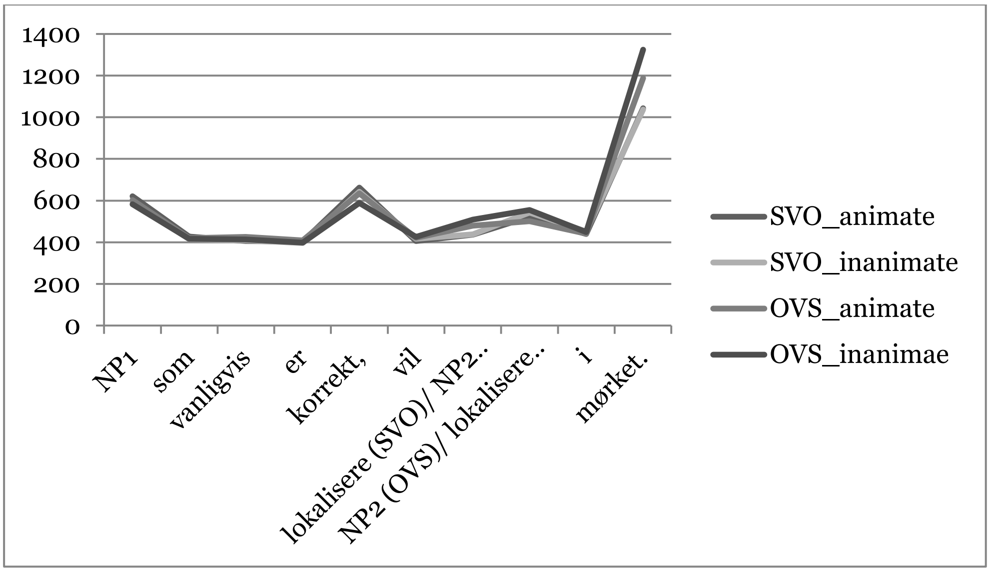

Figure B1.1 Mean reading times (in milliseconds) per word; L1 group

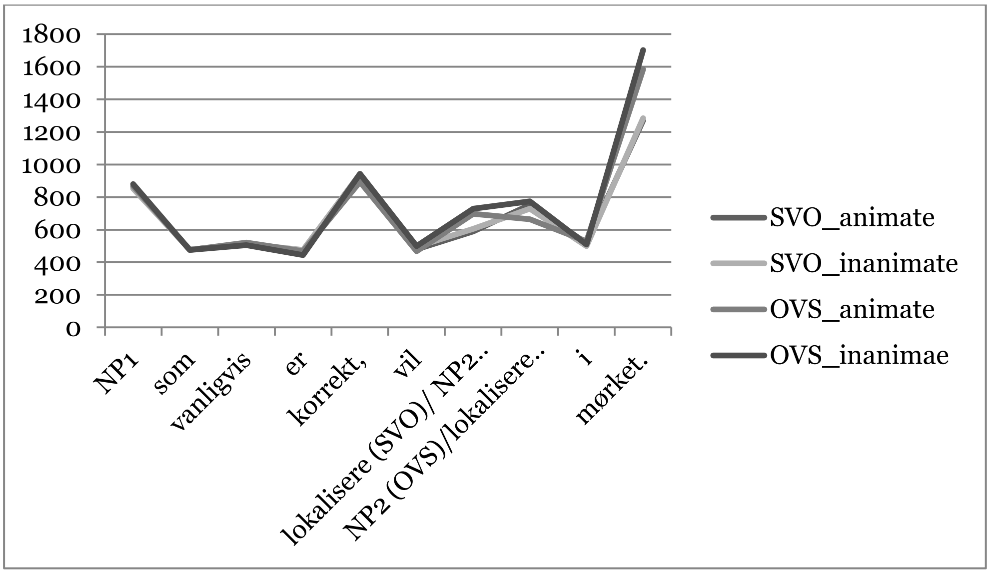

Figure B1.2 Mean reading times (in milliseconds) per word; L2 group

	SVO_animate	**SVO_inanimate**	**OVS_animate**	**OVS_inanimate**
Overall	1159 (443)	1170 (502)	1209 (440)	1320 (581)
L1	951 (268)	1000 (322)	1017 (246)	1098 (311)
L2	1381 (488)	1350 (594)	1413 (509)	157 (703)

Table B1.5 Overview of reading times (in msecs) by condition for both groups, SDs given in brackets

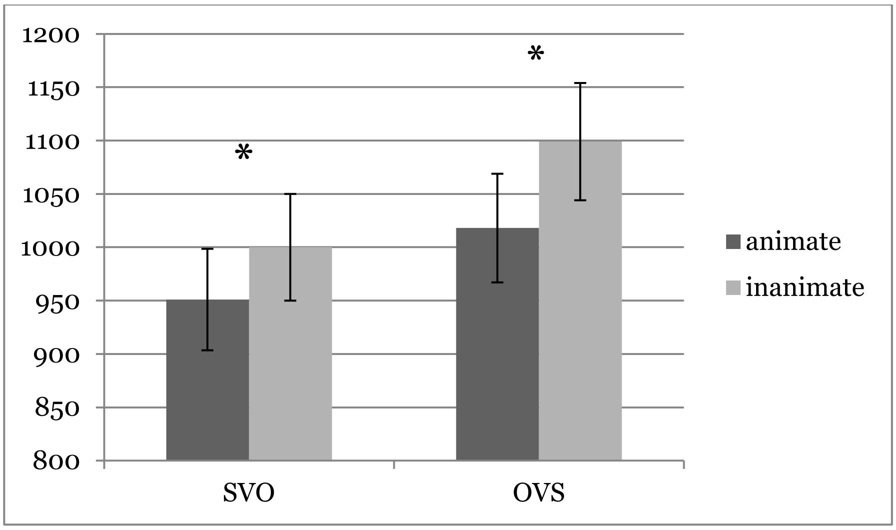

Figure B1.3 Reading times at region of manipulation; L1 group

An ANOVA run on the log-transformed reading times of the L1 group revealed an ME of Animacy in the by-subject analysis ($F_1(1,31)=5.61$, $p=0.02$, $F_2(1,23)=2.46$, $p=0.13$) and an ME of Order ($F_1(1,31)=11.89$, $p=0.002$, $F_2(1,23)=10.8$, $p=0.003$). There was no Animacy x Order interaction, Fs<1. Figure B1.3 shows the separate MEs with faster reading times for the SVO order and faster reading times for animate subjects.

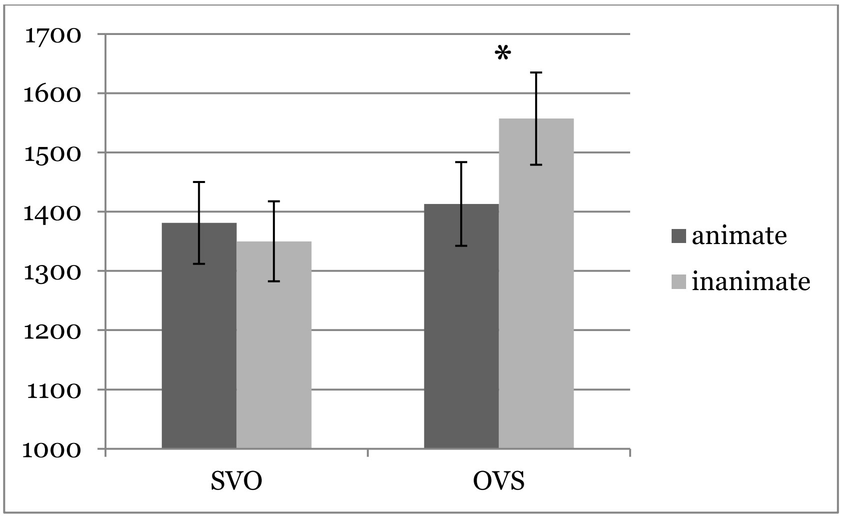

Figure B1.4 Reading times at region of measurement; L2 group

An ANOVA run on the log-transformed reading times of the L2 group revealed an ME of Order ($F_1(1,29)=10.37$, $p=0.003$, $F_2(1,23)=11.91$, $p=0.002$) and an Animacy x Order interaction in the by-subject ANOVA ($F_1(1,29)=6.72$, $p=0.015$, $F_2(1,23)=3.76$, $p=0.065$). There was no ME of Animacy, Fs<1. Figure B1.4

suggests that the interaction is caused by a presence of an effect of animacy in the OVS condition that is absent in the SVO condition. Post-hoc by-subject t-tests confirmed this visual trend: Animate subjects were read significantly faster than inanimate ones in the OVS condition ($t_1(29)= -2.22$, $p=0.035$) and there was no effect of animacy in the SVO condition ($t_1(29)= 1.01$, $p=0.3$).

Study 2 – German ditransitive sentences

Experiment 2a - Acceptability judgment task

(df=38)	AoA	Years learning German	Years living in German
Accusative-dative rating	t=-0.34 p=0.74	t=0.65 p=0.52	t=0.65 p=0.5

Table B2.1 Additional linear regressions with L2-specific data as predictors

Experiment 2b - SPR task

Case	Noun	Length	Typefrequency (log)	Lemmafrequency (log)
dative	Ehepaar	7	0.855	1.020
dative	Vater	5	2.338	2.436
dative	Zoo	3	0.563	0.602
dative	Publikum	8	1.620	1.707
dative	Regisseur	9	0.839	0.997
dative	Staatsanwalt	12	1.285	1.403
dative	Matrosen	8	0.764	0.936
dative	Investor	8	-0.582	0.302
dative	Kollegen	8	1.532	1.658
dative	Finanzamt	9	0.285	0.481
dative	Tierschutzverein	16	-0.625	-0.275
dative	Großvater	9	1.362	1.443
dative	Chorleiter	10	-0.324	-0.255
dative	Kritiker	8	1.088	1.177
dative	Hochzeitsplaner	15	-	-
dative	Torwart	7	-0.454	-0.415
accusative	Kaufvertrag	11	-0.066	0.168
accusative	Brief	5	1.989	2.219
accusative	Betrag	6	1.332	1.580
accusative	Roman	5	1.350	1.688
accusative	Walzer	6	0.485	0.598
accusative	Bericht	7	1.863	2.039
accusative	Befehl	6	1.665	1.827
accusative	Entwurf	7	1.532	1.700
accusative	Antrag	6	1.802	1.934
accusative	Ratschlag	9	0.013	0.957
accusative	Hauptgewinn	11	-0.389	-0.372
accusative	Artikel	7	1.925	2.010
accusative	Song	4	-0.148	0.570
accusative	Spielfilm	9	-0.083	0.267
accusative	Schritt	7	1.932	2.182
accusative	Hinweis	7	1.456	1.643

Table B2.2 Length, type and lemma frequencies for the object NPs

Segment	Extreme value cutoff	Removed data points	Percentage
1 (that)	lower: 150ms upper: 1500ms	30	2.84%
2 (subject article)	lower: 150ms upper: 1700ms	25	2.37%
3 (subject)	lower: 150ms upper: 2800ms	29	2.74%
4 (article object 1)	upper: 1700ms	31	2.93%
5 (object 1)	upper: 4500ms	34	3.22%
6 (article object 2)	upper: 2000ms	30	2.84%
7 (object 2)	upper: 3700ms	34	3.22%
8 (main verb)	upper: 3500ms	33	3.12%
9 (auxiliary)	upper: 2500ms	34	3.22%

Table B2.3 Data cleaning procedure for Experiment 2b

Segment	F1-ANOVA (1,66)	F2-ANOVA (1,15)
1 (that)	ME of Group F=16.00,p<0.001 ME of Order F=6.7, p=0.012	ME of Group F=183.3, p<0.001 ME of Order F=7.22, p=0.017
2 (subject article)	ME of Group F=14.62, p<0.001	ME of Group F=363.36, p<0.001
3 (subject)	ME of Group F=17.1, p<0.001	ME of Group F=430.97, p<0.001
4 (article object 1)	ME of Group F=15.85, p<0.001	ME of Group F=235.88, p<0.001
5 (object 1)	ME of Group F=31.76, p<0.001 ME of Order F=22.44, p<0.001 Group x Order F=5.9, p=0.018	ME of Group F=432.33, p<0.001 ME of Order F=9.7, p=0.007 Group x Order F=4.6, p=0.048
6 (article object 2)	ME of Group F=18.58, p<0.001 ME of Order F=5.42, p=0.023	ME of Group F=303.12, p<0.001
7 (object 2)	ME of Group F=40.09, p<0.001 ME of Order F=23.48, p<0.001	ME of Group F=369.61, p<0.001 ME of Order F=5.62, p=0.032
8 (main verb)	ME of Group F=46.5, p<0.001 ME of Order F=5.3, p=0.025	ME of Group F=747.18, p<0.001 ME of Order F=6.17, p=0.025
9 (auxiliary)	ME of Group F=23.05, p<0.001 ME of Order F=8.38, p=0.005	ME of Group F=250.4, p<0.001 ME of Order F=9.41, p=0.008

Table B2.4 Between-groups ANOVAs per segment for transformed RTS (reciprocal square root transformation), only significant or marginally significant values

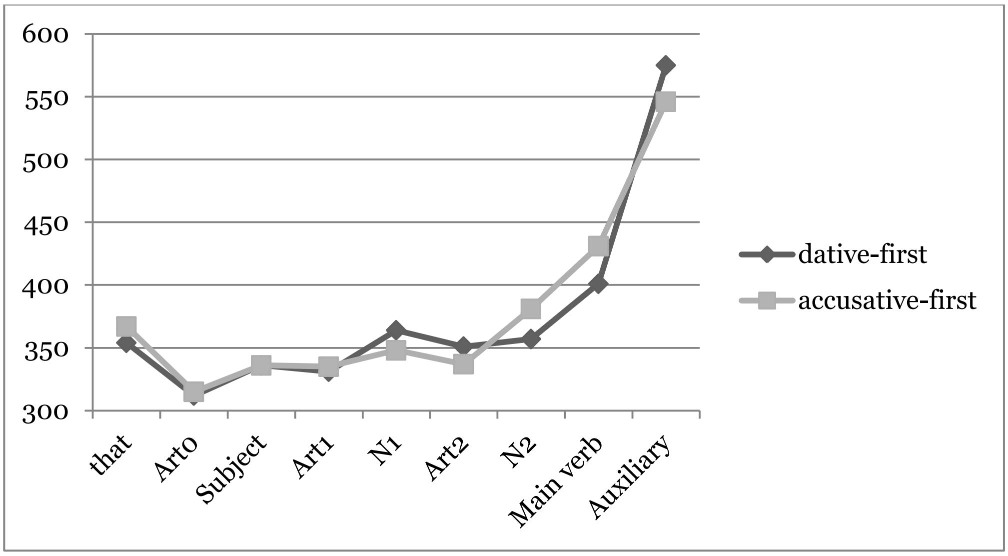

Figure B2.1 Reading times per segment; L1 group

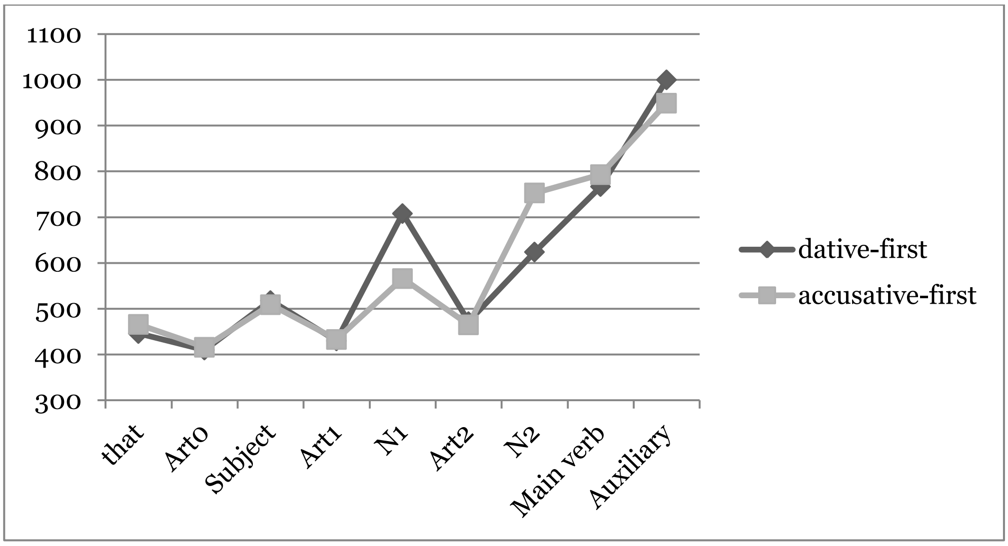

Figure B2.2 Reading times per segment; L2 group

(df=34)	Art1	N1	Art 2	N2	Main verb
Goethe	t=0.72 p=0.47	t= -0.4 p=0.69	t=0.83 p=0.41	t=0.92 p=0.37	t=1.75 p=0.089
AoA	t=0.24 p=0.81	t=0.74 p=0.47	t=0.79 p=0.44	t= -0.76 p=0.45	t= -1.28 p= 0.2
Years learning German	t=0.06 p=0.95	t=0.27 p=0.79	t= -1.16 p=0.25	t= -0.97 p=0.34	t=0.84 p=0.41
Years living in Germany	t=0.17 p=0.87	t=0.61 p=0.54	t= -1.33 p=0.19	t= -0.09 p= 0.92	t=1.67 p=0.1

Table B2.5 Linear regressions with L2-specific variables

Study 3: Norwegian particle verbs

Pilot study

Item	Object	Letters	Syllables	Rating (SD)
1	huset	6	2	3.31 (1.26)
	det berømte huset	15	6	2.67(1.18)
	det berømte gamle huset	20	8	2.78 (1.19)
2	buksen	6	2	1.7 (0.99)
	den varme buksen	14	5	2.52 (1.25)
	den pene varme buksen	18	7	3.08 (1.26)
3	søppelet	8	3	2.7 (1.3)
	det stinkende søppelet	20	7	2.73 (1.43)
	det ekle stinkende søppelet	24	9	2.3 (1.1)
4	båten	5	2	2.58 (1.36)
	den lille båten	13	5	2.78 (1.31)
	den enkle lille båten	18	7	3.37 (1.33)
5	fjernsynet	10	3	2.44 (1.01)
	det store fjernsynet	18	6	2.7 (1.27)
	det store brede fjernsyne	22	8	2.69 (1.26)
6	skjegget	9	2	2.37 (1.18)
	det lange skjegget	16	5	3.00 (1.36)
	det lange brune skjegget	21	7	3.11 (1.19)
7	hunden	6	2	1.42 (0.95)
	den blinde hunden	15	5	1.85 (0.95)
	den våte blinde hunden	19	7	1.89 (0.97)
8	passordet	9	3	2.93 (1.17)
	det kompliserte passordet	23	8	2.37 (1.15)
	det kompliserte sikre passordet	29	10	3.04 (1.18)
9	funnet	6	2	3.19 (1.3)
	det arkeologiske funnet	21	9	3.38 (1.44)
	det viktige arkeologiske funnet	28	12	3.89 (1.05)
10	søknaden	8	3	2.81 (1.3)
	den deprimerende søknaden	23	9	3.3 (1.2)
	den siste deprimerende søknaden	28	11	2.74 (1.23)
11	kaken	5	2	3.3 (1.17)
	den gigantiske kaken	18	7	2.7 (1.27)
	den gigantiske lekre kaken	23	9	3.12 (1.42)
12	ballongen	10	3	2.07 (0.96)
	den oransje ballongen	19	7	2.35 (1.29)
	den eneste oransje ballongen	25	10	2.81 (1.24)
13	programmet	10	3	2.31 (1.44)
	det aktualiserte programmet	26	10	3.33 (1.33)
	det aktualiserte bedre programmet	30	12	2.96 (1.34)

14	tanken	6	2	2.48 (1.25)
	den alvorlige tanken	18	7	2.41 (1.05)
	den intime alvorlige tanken	24	10	2.58 (1.24)
15	leketøyet	9	4	1.81(1)
	det fantastiske leketøyet	23	9	2.5 (1.39)
	det fantastiske nye leketøyet	26	11	2.74 (1.32)
16	bilen	5	2	1.65 (0.85)
	den ødelagte bilen	16	7	2.00 (1.07)
	den ødelagte billige bilen	23	10	2.3 (1.03)
17	ringen	6	2	2.3 (1.49)
	den gylne ringen	14	5	2.11 (1.22)
	den ekte gylne ringen	18	7	2.81 (1.27)
18	jakken	6	2	1.7 (0.99)
	den blå jakken	12	4	1.58 (0.95)
	den tjukke blå jakken	18	6	2.11 (0.93)
19	luften	6	2	2.31 (1.41)
	den arktiske luften	17	6	2.52 (1.05)
	den kalde arktiske luften	22	8	1.96 (0.85)
20	gaven	5	2	2.07 (1.07)
	den forferdelige gaven	20	8	2.59 (1.34)
	den forferdelige stygge gaven	26	10	2.65 (1.44)
21	romanen	7	3	2.89 (1.34)
	den spennende romanen	20	7	3.85 (1.16)
	den spennende korte romanen	24	9	4.00 (1.00)
22	hemmeligheten	13	5	4.23 (1.21)
	den fryktelige hemmeligheten	26	10	4.26 (0.81)
	den fryktelige sjokkerende hemmeligheten	37	14	4.15 (0.99)
23	rommet	6	2	3.78 (1.12)
	det mørke rommet	14	5	3.74 (1.13)
	det skitne mørke rommet	20	7	3.73 (1.08)
24	avtalen	7	3	3.74 (1.06)
	den overdrevne avtalen	20	8	3.88 (1.11)
	den overdrevne dyre avtalen	24	10	3.59 (1.15)

Table B3.1 Object length in letters and syllables and average rating

Experiment 3a - Acceptability judgment task

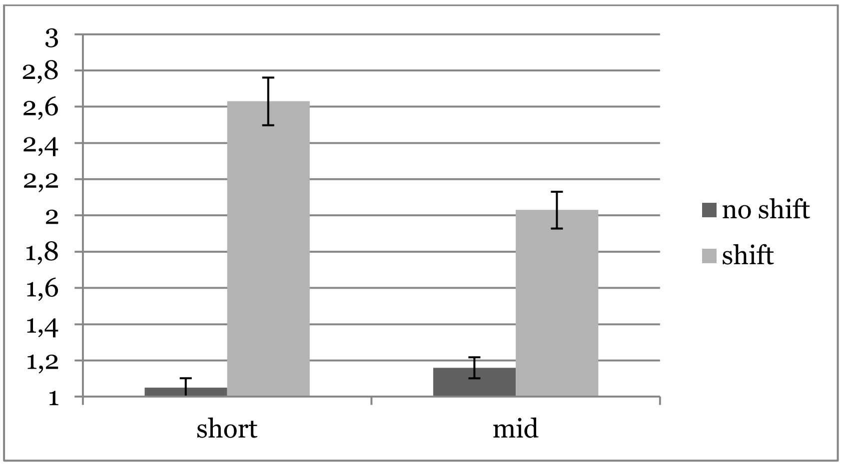

Figure B3.1 Average acceptability ratings; L1 group

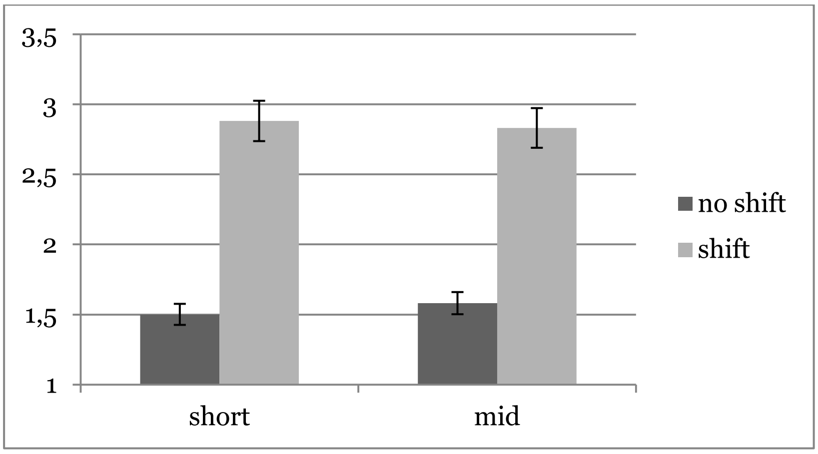

Figure B3.2 Average acceptability ratings; L2 group

Experiment 3b - SPR task

Segment	Extreme value cutoff	Removed data points	Percentage
1 (name)	upper: 3000ms	22	2.9%
2 (main verb)	lower: 140ms upper: 2500ms	24	3.16%
3 (object NP/particle)	lower: 140ms upper: 2500ms	19	2.5%
4 (particle/object NP)	upper: 2700ms	11	1.45%
5 (preposition)	lower: 150ms upper: 2500ms	25	3.29%
6 (NP)	lower: 150ms upper: 6000ms	17	2.24%

Table B3.2 Data cleaning procedure Experiment 3b; short condition

Segment	F1-ANOVA (1,62)	F2-ANOVA (1,23)
1 (name)	ME of Group F=8.63, p=0.005	ME of Group F=64.02, p<0.001

2 (main verb)	ME of Group F=9.95, p=0.002	ME of Group F=58.07, p<0.001
3 (object NP/particle)	ME of Group F=17.16, p<0.001 ME of Order F=75.33, p<0.001 Group x Order F=10.54, p=0.002	ME of Group F=102.74, p<0.001 ME of Order F=29.78, p<0.001 Group x Order F=10.08, p=0.004
4 (particle/object NP)	ME of Group F=16.2, p<0.001 ME of Order F=20.7, p<0.001 Group x Order F=15.54, p<0.001	ME of Group F=174.64, p<0.001 ME of Order F=16.74, p<0.001 Group x Order F=16.75, p<0.001
5 (preposition)	ME of Group F=8.73, p=0.004	ME of Group F=35.9, p<0.001
6 (NP)	ME of Group F=13.65, p<0.001	ME of Group F=90.49, p<0.001

Table B3.3 Between-groups ANOVAs per segment for log-transformed RTs, only significant or marginally significant values, short condition

Segment	**Extreme value cutoff**	**Removed data points**	**Percentage**
1 (name)	lower: 150ms upper: 3000ms	24	3.18%
2 (main verb)	lower: 150ms upper: 2000ms	21	2.76%
3 (article/particle)	upper: 3000ms	21	2.76%
4 (adjective/article)	upper: 3000ms	19	2.52%
5 (adjective/adjective)	upper: 3000ms	18	2.39%
6 (object NP/adjective)	lower: 150ms upper: 3000ms	24	3.18%
7 (particle/object NP)	upper: 3000ms	19	2.52%
8 (preposition)	upper: 1400ms	17	2.25%
9 (NP)	upper: 12000ms	18	2.39%

Table B3.4 Data cleaning procedure Experiment 3b; long condition

Segment	**F1-ANOVA (1,62)**	**F2-ANOVA (1,23)**
1 (name)	ME of Group F=15.67, p<0.001	ME of Group F=77.69, p<0.001
2 (main verb)	ME of Group F=12.44, p<0.001 Group x Order F=6.06, p=0.017	ME of Group F=35.9, p<0.001
3 (article/particle)	ME of Group F=4.36, p=0.041 Group x Order F=4.98, p=0.03	ME of Group F=33.0, p<0.001
4 (adjective/article)	ME of Group F=13.4, p<0.001 ME of Order	ME of Group F=103.24, p<0.001 ME of Order

	F=101.88, p<0.001 Group x Order F=4.7, p=0.034	F=91.89, p<0.001 Group x Order F=13.99, p=0.001
5 (adjective/adjective)	ME of Group F=17.32, p<0.001 ME of Order F=16.72, p<0.001 Group x Order F=13.4, p<0.001	ME of Group F=147.68, p<0.001 ME of Order F=12.32, p=0.002 Group x Order F=8.02, p=0.009
6 (object NP/adjective)	ME of Group F=17.17, p<0.001 Group x Order F=5.37, p=0.024	ME of Group F=84.2, p<0.001
7 (particle/object NP)	ME of Group F=13.27, p<0.001 ME of Order F=48.99, p<0.001 Group x Order F=7.2, p=0.009	ME of Group F=78.53, p<0.001 ME of Order F=22.21, p<0.001 Group x Order F=8.46, p=0.008
8 (preposition)	ME of Group F=4.6, p=0.036 Order *F=3.3, p=0.07*	ME of Group F=15.61, p<0.001
9 (NP)	ME of Group F=5.38, p=0.024	ME of Group F=31.93, p<0.001

Table B3.5 Between-groups ANOVAs per segment for log-transformed RTs, only significant or marginally significant values, long condition

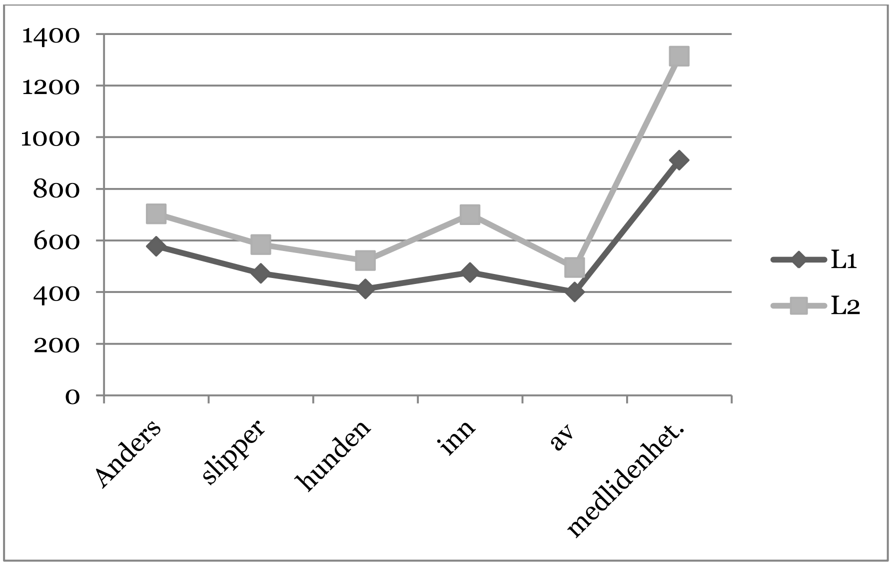

Figure B3.3 Reading times for the OP order, short condition

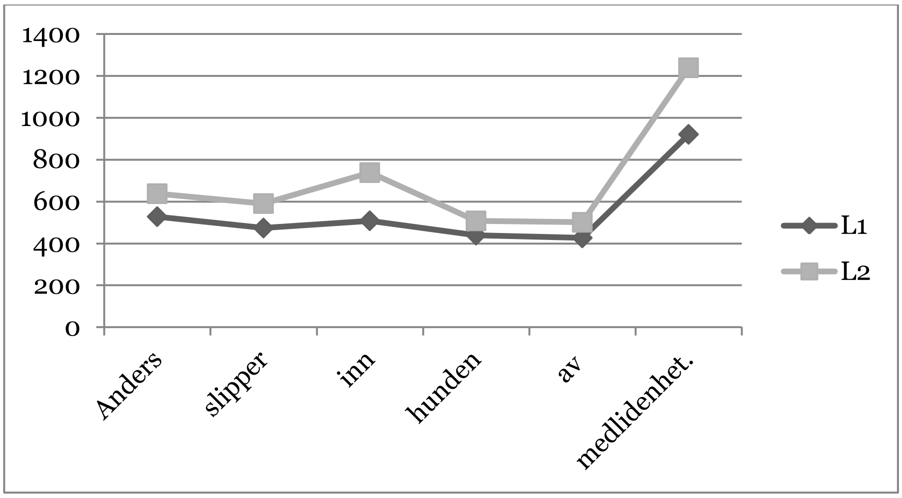

Figure B3.4 Reading times for the PO order, short condition

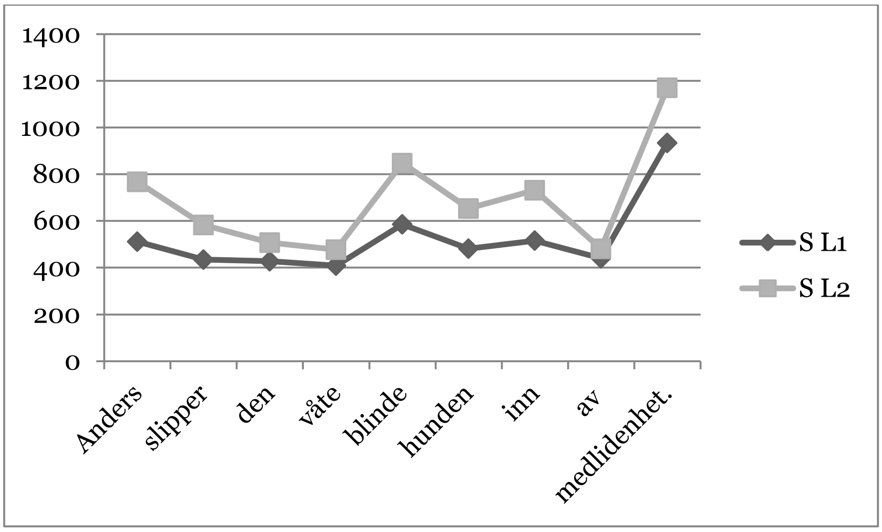

Figure B3.5 Reading times for the OP order, long condition

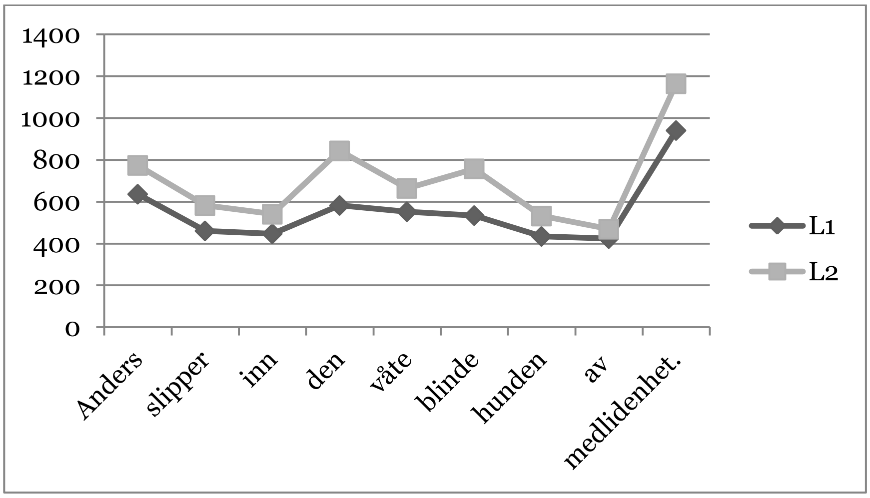

Figure B3.6 Reading times for the PO order, long condition

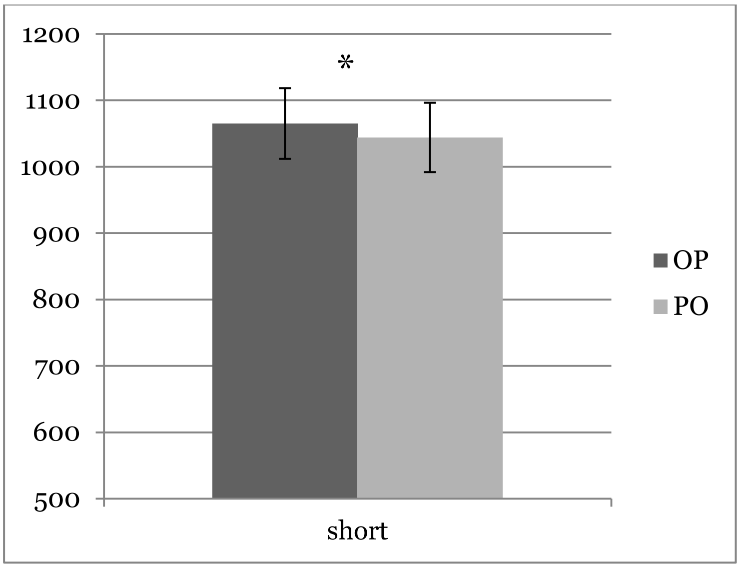

Figure B3.7 Reading times of the ROM; both participant groups, short condition

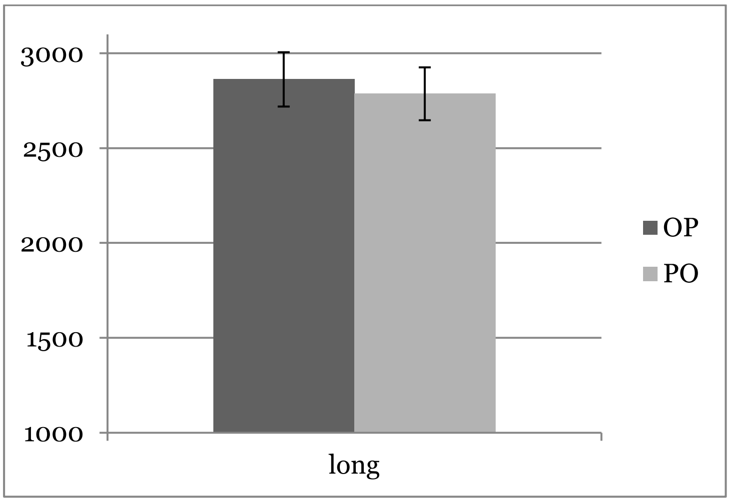

Figure B3.8 Reading times of the ROM; both participant groups, long condition

Paired t-tests showed a significant ME of Order in the short condition ($t_1(63)$= 2.74, p=0.008, $t_2(23)$= 2.02, p=0.055). There was no ME of Order in the long condition ($t_1(63)$= 1.63, p=0.11, $t_2(23)$= 1.15, p=0.26).

(df=30)	**Self-rating**	**AoA**	**Years learning Norwegian**	**Months living in Norway**
Short condition	t=0.34 p=0.74	t=0.64 p=0.52	t=-1.14 p=0.267	t=-0.65 p=0.52
Long condition	t=0.19 p=0.85	t=0.34 p=0.74	t=-0.38 p=0.71	t=0.6 p=0.55

Table B3.9 Linear regressions with L2-specific variables

Study 4 – German particle verbs

Experiment 4a - Acceptability rating task

(df=37)	**Goethe score**	**AoA**	**Years learning German**	**Years living in Germany**
Accuracy scores	t=19.3 p=0.06	t=-1.54 p=0.133	t=1.38 p=0.18	t=0.12 p=0.9

Table B4.1 Linear regressions with L2-specific variables

Experiment 4b – SPR task

Segment	**NAs from residualization**	**Removed data points (+/- 2.5 SD)**	**Percentage (1513 = 100%)**
Precritical region 4	11	28	2.58%
Precritical region 3	1	23	1.58%
Precritical region 2	7	30	2.45%
Precritical region 1	7	26	2.18%
Matrix verb	22	32	3.57%
Particle/Spillover	25	43	4.49%

Table B4.2 Data cleaning for Experiment 4b

Segment	**F1-ANOVA (1,62)**	**F2-ANOVA (1,23)**
Precritical region 4	ME of Group F=38.33, p<0.001 ME of Type F=38.67, p<0.001	ME of Group F=156.32, p<0.001 ME of Type F=37.85, p<0.001
Precritical region 3	ME of Group F=39.84, p<0.001 ME of Split F=4.16, p=0.046	ME of Group F=87.29, p<0.001 ME of Split F=5.96, p=0.023
Precritical region 2	ME of Group F=17.02, p<0.001	ME of Group F=10.74, p=0.003
Precritical region 1	ME of Type F=20.00, p<0.001 Group x Type F=7.91, p=0.007	ME of Type F=15.94, p<0.001 Group x Type F=6.27, p=0.02
Matrix verb	ME of Group F=13.94, p<0.001 Group x Split F=15.03, p<0.001	ME of Group F=29.07, p<0.001 Group x Split F=15.73, p<0.001
Particle/Spillover	ME of Group F=11.4, p=0.001 ME of Type *F=3.03, p=0.087*	ME of Group F= 17.73, p<0.001

Table B4.5 Between-groups ANOVAs per segment for transformed RTS (residual reading times), only significant or marginally significant values

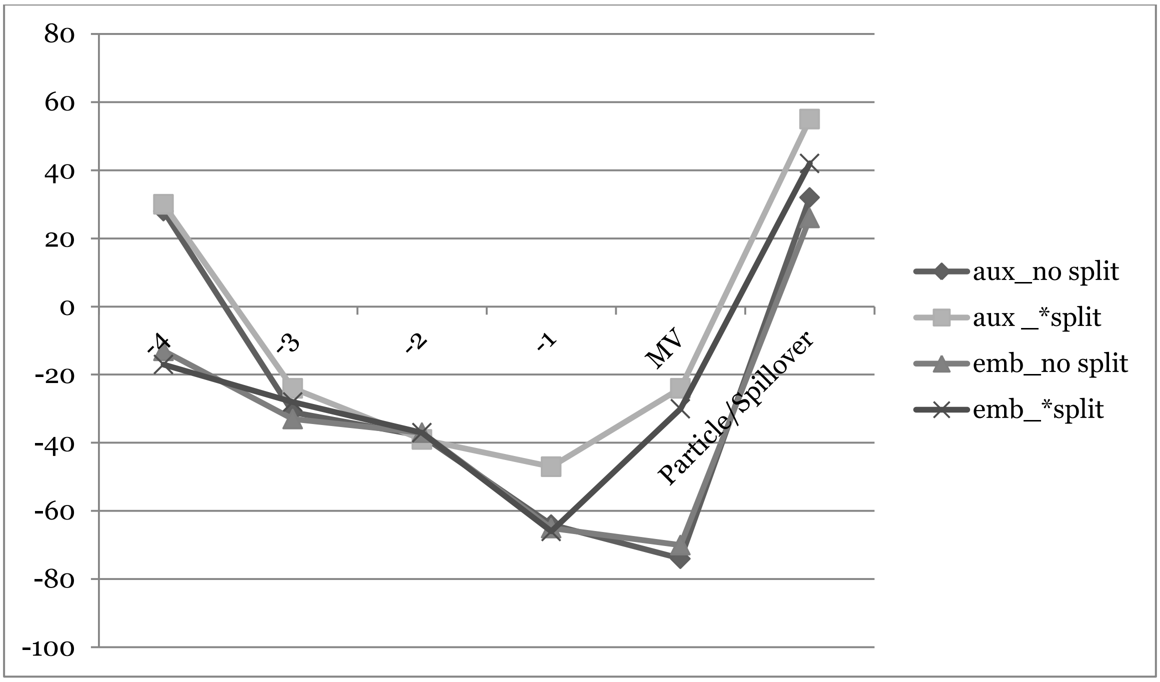

Figure B4.1 Residual reading times for core dataset, L1 group

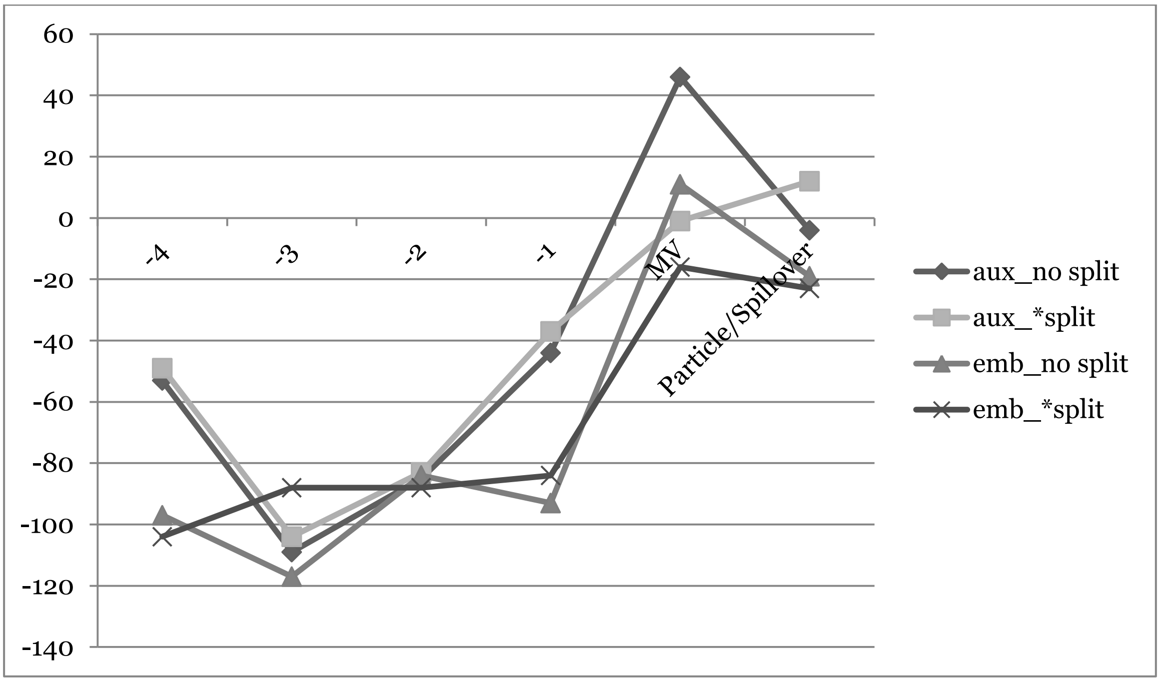

Figure B4.2 Residual reading times for core dataset, L2 group

(df=31)	**Goethe score**	**AoA**	**Years learning German**	**Years living in Germany**
Main verb	t=-2.73 p=0.01	t=2.02 p=0.052	t=-1.89 p=0.068	t=-2.85 p=0.008
Particle/spillover regio	t=-0.5 p=0.063	t=-1.55 p=0.09	t=0.22 p=0.83	t=-1.2 p=0.24

Table B4.5 Linear regressions with L2-specific variables

APPENDIX C – Participants

Study 1 and 3 (Norwegian)

ID	Age	Sex	Handedness	Other information
N101	30	male	right	
N102	20	female	right	juvenile epilepsy
N103	20	female	left	
N104	33	male	left	
N105	38	female	right	excluded from experiment 1a
N106	44	male	ambidextrous	
N107	41	female	right	colorblind
N108	34	male	right	
N109	38	female	right	
N110	25	female	right	
N111	20	female	right	
N112	18	female	right	international adoptee with L1 Korean
N113	21	female	right	
N114	24	female	right	
N115	34	male	right	excluded from experiment 1a
N116	21	female	left	
N117	20	female	right	
N118	25	male	right	early L2 English
N119	20	female	right	
N120	27	female	left	second L1 Spanish
N121	19	female	right	

N122	44	female	right	
N123	61	female	right	retrobulbary neuritis, Morton's neuroma
N124	46	female	right	
N125	19	female	right	
N126	22	female	right	
N127	38	female	right	
N128	50	female	right	
N129	21	male	right	
N130	22	female	right	
N131	21	male	right	
N132	23	male	right	

Table C1.1 Biographical data of L1 Norwegian participants

ID	**Age**	**Sex**	**Handedness**	**Exclusion criterion**
N133	31	male	right	long reading times, low accuracy in offline task (general exclusion)

Table C1.2 Biographical data of excluded L1 Norwegian participants

ID	**Age**	**Sex**	**Handed-ness**	**LX NOR**	**AoA NOR**	**Time learning NOR (years)**	**Stay in NOR (months)**	**Rating overall**	**Rating reading**	**Rating writing**	**Rating listening**	**Rating speaking**	**Other information**
N201	27	female	right	4	19	8	13.5	22	6	6	5	5	
N202	23	female	right	5	19	1	2	16	5	4	4	3	tinnitus
N203	24	male	right	4	20	4	6	14	4	3	4	3	
N204	29	male	right	4	26	2	24	17	4	4	5	4	
N205	28	male	right	4	19	9	6	17	5	4	4	4	

N206	22	male	right	4	19	4	12	17	5	4	4	4	
N207	26	female	right	5	21	5	6	15	4	4	4	3	
N208	27	female	right	4	17	8	12	15	5	3	4	3	additional knowledge of Icelandic
N209	58	male	right	4	18	9.5	18	18	6	4	5	3	
N210	21	female	right	6	20	0.5	0	17	5	4	4	4	additional knowledge of Swedish, Danish and Icelandic
N211	22	male	right	5	21	4	7	22	6	5	6	5	
N212	22	female	left	4	18	4	12	20	5	4	6	5	additional knowledge of Icelandic
N213	39	male	right	10	29	10	36	15	5	2	4	4	additional knowledge of Swedish, Icelandic and Danish, living in Norway
N214	25	female	right	4	23	2	4	21	6	5	5	5	
N215	27	female	right	6	24	3	18	13	4	2	3	4	living in Norway
N216	23	female	right	4	14	9	84	24	6	6	6	6	
N217	23	male	right	4	17	5	3	14	4	3	4	3	
N218	22	female	right	5	19	3.5	12	15	5	4	4	2	
N219	20	male	right	4	16	1	12	16	5	3	5	3	
N220	25	female	right	4	23	2.5	8	16	4	4	4	4	excluded from Experiment 1

													due to low offline accuracy
N221	23	female	right	6	21	2	0	16	5	4	3	4	
N222	30	male	right	5	24	1	12	18	5	5	4	4	
N223	23	female	right	7	18	5	18	16	3	5	4	4	additional knowledge of Swedish and Icelandic
N224	30	male	right	5	19	3	12	22	6	5	5	6	additional knowledge of Swedish and Icelandic, excluded from Experiment 1a due to low offline accuracy
N225	22	female	right	5	19	3	0	18	5	4	4	5	
N226	23	female	right	4	21	2	0	17	5	4	4	4	
N227	29	female	right	5	19	10	60	18	5	4	5	4	living in Norway
N228	26	male	right	3	20	6	6	13	5	3	3	2	
N229	25	male	right	4	22	3	12	16	5	4	4	3	
N230	24	male	right	4	19	5	10	17	5	4	4	4	
N231	27	female	left	5	15	12	72	23	6	5	6	6	living in Norway
N232	23	female	right	4	19	3	12	16	5	4	3	4	

Table C1.3 Biographical data of L2 Norwegian participants

Study 2 and 4 (German)

ID	Age	Sex	Handedness	Goethe score	Other information
G101	27	female	right	28	
G102	39	male	right	29	
G103	24	female	right	28	
G104	21	female	right	29	
G105	26	male	left	29	
G106	21	female	right	27	
G107	23	female	right	29	
G108	20	female	right	28	
G109	19	female	left	30	
G110	24	female	right	30	migraine
G111	26	female	right	27	
G112	18	female	right	28	
G113	23	female	right	28	
G114	22	female	right	29	
G115	19	female	right	29	
G116	27	female	right	29	
G117	24	female	right	25	
G118	29	female	right	28	epilepsy
G119	23	female	right	27	
G120 (=N206)	23	male	right	29	
G121 (=N218)	23	female	right	30	
G122 (=N203)	25	male	right	29	

G123 (=N228)	27	male	right	29	
G124	21	female	right	30	
G125	23	female	right	28	
G126	25	female	right	29	
G127	20	female	right	29	anosmia
G128	26	male	right	29	
G130	26	male	right	27	
G131	23	female	right	27	
G132	20	female	right	29	

Table C2.1 Biographical data of L1 German participants

ID	**Age**	**Sex**	**Handed-ness**	**Goethe score**	**Exclusion criterion**
G129	27	female	right	29	low accuracy in online task (excluded from Experiment 2b and 4b)
G133	24	female	left	28	low accuracy in online task (excluded from Experiment 2b and 4b)

Table C2.2 Biographical data of excluded L1 German participants

ID	**Age**	**Sex**	**Handed-ness**	**L1**	**LX GER**	**AoA GER**	**Time learning GER (years)**	**Age of Arrival in GER**	**Stay in GER (years)**	**Goethe score**	**Other information**
G202	25	f	right	Russian	3	12	13	25	0.5	21	
G204	34	m	right	Serbian	3	32	2	32	2	24	

G205	34	f	right	Russian	3	18	16	23	11	24	
G206	26	f	right	Ukrainian/Russian	4	19	4	21	5	23	
G207	21	f	right	Russian	2	5	16	5	16	30	
G208	23	f	right	Russian	3	15	8	19	4	23	
G209	24	m	right	Polish	3	17	5	22	2	24	
G210	23	f	right	Ukrainian/Russian	4	13	8	22	1	19	
G211	24	f	left	Czech	3	14	10	24	0.5	26	
G212	25	f	right	Polish	3	13	12	22	3	27	squinting
G213	28	f	right	Russian	5	21	7	21	7	27	
G214	21	f	right	Polish	2	15	6	15	6	23	
G215	22	f	right	Czech	3	12	8	22	0	16	
G216	31	f	right	Russian	2	7	15	31	0	30	
G218	24	m	right	Slovenian	2	7	17	24	0.6	25	
G219	33	f	right	Bulgarian	4	15	18	20	13	27	
G220	23	f	right	Czech/Slovak	4	12	11	17	6	28	
G221	26	f	right	Czech	3	15	11	25	1	19	squinting
G222	27	m	left	Czech	2	10	17	25	2	26	
G223	20	f	right	Russian	2	7	13	7	13	27	
G224	24	f	right	Bulgarian	2	13	5	24	0.5	24	
G226	27	f	right	Russian	2	15	12	26	1	21	
G229	22	f	right	Russian	2	7	15	8	14	28	
G230	30	f	right	Russian	3	7	23	20	10	27	
G231	32	m	right	Russian	3	14	18	18	14	25	
G232	26	f	right	Czech	3	16	5	25	1	12	
G233	22	m	right	Russian	2	10	12	10	12	26	
G234	23	f	left	Russian	3	12	11	12	11	28	
G235	27	m	right	Russian	3	5	22	5	22	28	

G237	29	f	right	Polish	3	15	14	20	9	24	
G238	25	f	right	Polish	3	13	12	18	7	26	
G239	29	f	right	Macedonian	3	28	1.5	28	1	19	
G240	22	f	right	Slovenian	3	12	10	22	0.5	27	

Table C2.3 Biographical data of L2 German participants

ID	**Age**	**Sex**	**Handed-ness**	**L1**	**LX GER**	**AoA GER**	**Time learning GER (years)**	**Age of Arrival GER**	**Stay in GER (years)**	**Goethe score**	**Exclusion criterion**
G201	21	f	right	Russian	2	8	13	21	0	25	low accuracy in offline task (Experiment 4)
G203	30	f	left	Russian	3	29	2	29	1	18	low accuracy in offline task (Experiment 4)
G217	21	f	right	Russian	2	7	14	20	1	19	low accuracy in online task (Experiment 2b and 4b)
G225	24	f	right	Polish	3	16	8	22	2	21	low accuracy in online task (Experiment 2b and 4b)
G227	25	m	right	Russian/ Ukrainian	4	21	4	20	5	18	low accuracy in offline task (Experiment 4)
G228	25	f	right	Polish	1	0	25	22	3	29	bilingual from birth
G236	30	f	right	Russian	3	20	10	20	10	24	low accuracy in online task (Experiment 2b and 4b)
G241	26	f	right	Montenegrin	2	3	23	0	26	29	heritage speaker

Table C2.4 Biographical data of excluded L2 German participants

Erklärung

Gemäß §4 (2) 4. und §4 (2) 7. der Promotionsordnung der Humanwissenschaftlichen Fakultät der Universität Potsdam vom 15.05.2013.

Hiermit erkläre ich, Silke Schunack, 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.

Eltville, den 11.7.2016