6. The Morphosyntax of Scrambling at L2 Ultimate Attainment

Introduction

In this chapter, I consider the syntax-morphology interface in L2 off-line and on-line performance in the context of German scrambling. Three experiments test whether L2ers converge on the morphosyntax of scrambling in off-line judgements (Experiment 1), whether L2ers effect incremental syntactic reanalysis according to morphological cues in on-line reading (Experiment 2), and whether L2ers show native patterns of reanalysis in speeded judgements (Experiment 3).

The results show that near-native L2ers across L1s converge on target patterns of incremental reanalysis across the three experiments. However, the degree of target-like performance among the L2 groups varies according to (a) proficiency level, (b) L1 and (c) task. I argue that these effects can be systematically related to the computational load of L2 processing. I conclude that the patterns of (non-)convergence on the morphosyntax of scrambling obtained in Experiments 1-3 are compatible with the Fundamental Identity Hypothesis, i.e. that non-native and native grammars and processors are qualitatively identical.

This chapter is structured as follows: First, Section 6.1 gives an overview of the experiments and presents the research questions. Second, Section 6.2 presents background on the participants; then, Section 6.3 outlines the proficiency testing that was used to allocate the participants to proficiency groups. Subsequently, I present the rationale and discuss the results from each of the three experiments: Section 6.4 reports on off-line judgements on the morphosyntax of scrambling (Experiment 1). Section 6.5 recaps the essentials of the native processing of scrambling, which form the background for Experiments 2 and 3. Experiment 2 is reported in Section 6.6, and Experiment 3 in Section 6.7. In the general discussion (Section 6.8), I put the results from Experiments 1-3 together in order to characterize the processing of morphosyntax in advanced to near-native L2 speakers, and I relate it to the Fundamental Identity Hypothesis.

6.1. Overview and research questions

This chapter presents and discusses three experiments that test whether L2 speakers converge on target patterns of morphosyntax in off-line and on-line behaviour.

Previous research on L2 ultimate attainment finds that L2ers have difficulties with L2 inflectional morphology (Chapter 1, Franceschina, 2005; Lardiere, 2007; White, 2003a). These difficulties are modulated by L1 influence (Franceschina, 2005). Analogously, (non-near-native) L2ers fail to use morphosyntactic cues to native degrees in L2 processing (Jackson, 2005; Jiang, 2004; Sabourin, 2003). Against the background of these findings, one major question in current research is whether non-convergence in
off- and on-line performance on morphosyntax is indicative of representational impairment in L2 grammars (e.g. Clahsen & Felser, 2006b; Hawkins & Chan, 1997; Tsimpli, 2003) or computational problems in L2 use and comprehension (McDonald, 2006; Prévost & White, 2000b).

Since German OS orders derived by scrambling are morphologically marked by case, convergence on scrambling in the L2 requires target use of case morphology. An off-line judgement task (Experiment 1) tests whether advanced to near-native L2 speakers converge on grammatical knowledge of the syntax of scrambling and the morphological means of indexing word order alternations by case marking. Two on-line experiments examine whether morphosyntax is used in the real-time processing of scrambling. Specifically, a self-paced reading task probes whether L2ers show incremental effects of reanalysis in scrambling (Experiment 2). The results tell us whether L2ers establish syntactic dependencies on-line by using morphological cues. Second, a speeded grammaticality judgement task considers whether reanalysis is differentially informed by morphological feature type (Experiment 3). Its results tell us whether the grammatical informativity of morphosyntactic features is implicated in L2 processing. Accordingly, the three general questions addressed in three experiments are, respectively:

(Q1) Do L2 speakers show knowledge of the morphosyntax of scrambling in off-line judgements?
(Q2) Do L2 speakers show incremental reanalysis in on-line reading?
(Q3) Do L2 speakers use morphological information for reanalysis in speeded on-line judgements?

The three experiments are each conducted with 20 L1 English, 20 L1 Dutch and 19 L1 Russian learners, so that comparisons between different L1 groups can be made. Participants were allocated to two proficiency groups (see below), matched across L1s, so that comparisons based on proficiency level could be carried out. This way, we can differentiate between effects of proficiency and L1 effects. Finally, each participant completed all three experiments, so that within-group comparisons across tasks are possible. This means that three additional comparative research questions are formulated:

(Q4) Are there differences between off-line judgements and on-line processing of scrambling?
(Q5) Are there L1 differences in off-line judgements or on-line processing of scrambling depending on whether the L1 instantiates scrambling (English versus Dutch and Russian) or depending on whether the L1 instantiates scrambling and case marking (English and Dutch versus Russian)?
(Q6) Are there proficiency differences in judgements or processing of scrambling constructions?
Table 6.1, adapted from Chapter 3, summarizes the varying L1 properties of the groups in the morphosyntax of word order variation.

<table>
<thead>
<tr>
<th></th>
<th>Target</th>
<th>L1s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GERMAN</td>
<td>ENGLISH</td>
</tr>
<tr>
<td>Basic word order</td>
<td>SOV</td>
<td>SVO</td>
</tr>
<tr>
<td>Short scrambling</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Medium scrambling</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Syntax-Morphology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Case)</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Syntax-Morphology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Verbal agreement)</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 6.1. Cross-linguistic differences in the morphosyntax of scrambling.

The cross-linguistic differences in word order and case marking implicate different conditions for L1 transfer. English does not instantiate scrambling or case marking on full NPs and it has SVO in embedded clauses. Dutch parallels English in its lack of case morphology on full NPs; however, in basic word order (SOV) and limited scrambling, Dutch is like German. Finally, Russian has scrambling and overt case marking, yet basic word order is SVO. Experiments 1-3 will ascertain to what extent the differences between the L1s in basic word order, scrambling or case morphology affect (non-)convergence on scrambling in L2 German.

6.2. Participants

For the experiments reported in this chapter, 20 advanced to near-native L1 English speakers (13 female; 2 left-handed), 19 L1 Russian speakers (17 female; 0 left-handed), 20 L1 Dutch speakers (15 female; 2 left-handed) and 20 native German controls (9 female; 3 left-handed) were recruited. All participants had normal or corrected-to-normal vision and were apparently healthy at the time of testing. To make sure that all non-native speakers in this study clearly were adult L2 learners, only participants who started learning German after age 11 were included in this study. All non-native participants were residents in Germany at the time of testing, and they had each been exposed to German for more than ten years. All non-natives had attended higher education. Their ages ranged from 21-66 years (see Table 6.2). Participation in the experiments was exclusively contingent on age of onset, residence in Germany and the results on the proficiency test. In order to allow for a larger set of advanced to near-native L2 speakers to be tested in the study, there was no criterion of minimum length of residence or minimum length of exposure. In fact, all participants had spent more than three years in German-speaking countries and had had extensive exposure to German. All participants completed a questionnaire about their language backgrounds (see Appendix A for additional background characteristics of the participants). The non-native participants
were recruited through personal networks and postings in on-line forums and
newsgroups, and through English-, Russian-, or Dutch-speaking clubs, societies,
churches, kindergardens and schools. Since the recruitment materials stressed the
importance of advanced proficiency in German, participants who volunteered to take part
in the study tend to represent a self-selected, highly proficient subsample of late L2
learners. The non-native participants received chocolates and took part in a prize draw of
book vouchers as reimbursement for participation.

6.3. Proficiency measure and allocation to proficiency groups

In order to estimate the proficiency in German of the non-native participants, two tasks
on comprehension and production skills in the target language were administered: a C-
test and a picture-description task eliciting spontaneous speech.

6.3.1. C-test

A C-test served as the main independent measure of proficiency in this study. C-tests are
written tests of general language proficiency that present several texts with missing
information (for details, see, e.g., Grotjahn, 2002). A C-test consists of a number of texts:
Starting from the second sentence of each text, the second half of every second word is
deleted, and participants have to fill in the missing endings of words (1).

(1) This is an example of a C-test. Starting fr__ the sec___ sentence o__ each te__,
the sec___ half o__ every …

In correlational studies, C-tests have been shown to yield a better fit with extensive
proficiency tests than, e.g., cloze tests (e.g. Grotjahn, 2002 and references therein). For
this study, a C-test was chosen mainly for three reasons: (a) A C-test is made up of
several authentic texts from different topics and genres and thus does not advantage
participants with specialist lexical knowledge of a particular subject matter. (b) A C-test
focuses less on lexical knowledge than the cloze test does and more on grammatical
knowledge. Since the main experiments in this study concern (the use of) grammatical
knowledge, it seemed favourable to employ a proficiency measure that taps these aspects.
(c) A C-test allows for the exact scoring of responses without loss of accuracy since semi-
deletions of words generally do not permit acceptable alternative completions.

A 113-item C-test consisting of five authentic texts taken from newspapers,
magazines and book chapters was constructed (see Appendix B). The individual texts
contained between 16 and 27 deletions and ranged in subject matter from environmental
policy via culture to nuclear theory. The test was piloted on several advanced non-native
speakers of German, none of whom took part in the main study. Following some minor
changes, the test was administered on the Internet, where participants could log onto the website. Instructions preceded the task, and participants were told to complete the test in one session. There was a time limit of three minutes for each text. Several safeguards in the programming ensured that participants (a) did the test in one session, (b) could not go back to or redo (parts of) the texts for which the time limit had elapsed and (c) could not do the test more than once. Scoring was automatic and only exact matches were counted as correct. The combined accuracy score for all texts was recorded and also displayed for the participants upon completion of the task. Table 6.2 shows that native speakers scored between 76% and 93%, and the non-natives scored between 43% and 89%.

Given the comparatively wide score variation among the non-natives, they were subdivided into proficiency groups in order to be able to consider potential effects of proficiency level in the main experiments. The fairly small number of participants in each language group did not allow for numerous proficiency groups; the non-natives were therefore allocated to two proficiency groups according to the median score in the cloze test for each language group, namely, advanced (<67%) and near-native (≥67%). This led to a very balanced distribution of participants in groups for each language. Note that the near-natives did not all fall into the native range of C-test scores, although there was some overlap with 3 L1 English, 4 L1 Dutch and 2 L1 Russian subjects scoring within the native range.

A one-way ANOVA with Proficiency score as the dependent variable and Group (advanced English, near-native English, advanced Dutch, etc.) as the independent variable showed a significant effect of group (F(5,52) = 28.379, p<0.001). Post-hoc pairwise t-tests for independent samples (Bonferroni adjustment) on the proficiency scores demonstrate that there are no statistically significant differences either between any of the three advanced groups (p=1) or between any of the near-native groups (p≥0.938), yet there are statistically significant differences between the advanced groups, on the one hand, and the near-native groups, on the other hand (p<0.001). For proficiency groups collapsed across L1s, the near-natives are significantly different from the advanced group (F(1,56) = 11.676, p<0.001). These comparisons underline the differences in proficiency between the advanced and near-native groups. Detailed participant information is listed in Table 6.2.
On top of completing the C-test, all non-native participants and some native speakers took part in a speech-elicitation task based on a picture description. This task was selected in order to gain a glimpse of the production of the L2 and aspects, such as accent, fluency, etc., that cannot be tapped in a standardized comprehension-based test.

In this task, participants were asked to describe and talk about a picture and a cartoon-strip consisting of four frames. The picture was a copy of the “Cookie Theft” picture from the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983), and the cartoon strip depicted a tragic and violent event at a railway crossing. The speech samples of the subjects were digitally recorded and stored on a computer. Participants took between 45 seconds and 5 minutes 30 seconds for this task. The speech samples were analysed in two ways: first, an index of errors/mistakes was computed for each participant; second, each speech sample was rated by 3 native speakers of German. The index gives a picture of the accuracy in the L2 production of morphosyntax and the TL vocabulary, and the rating gives an estimate of the degree of native-likeness of the L2 participants in speech production.

The results from the two analyses are presented here to support the allocation of participants to proficiency groups according to the C-test. It was decided not to use the speech-elicitation data for dividing participants according to proficiency level because of the non-standardized nature of the task and analysis. Four reasons were relevant: (a) There is drastic variation in the length of individual speech samples. (b) Subjects could use avoidance strategies in their production to eschew specific grammatical or lexical
knowledge deficits. (c) Calculating an error index is hampered by the syncretism in, e.g.,
the German article system, which potentially leads to an underestimation of the actual
errors/mistakes in production. (d) Even with trained judges, rating speech samples is
inherently a non-standardized way of gauging proficiency. Therefore, the findings of the
speech-elicitation data will be presented to provide supplementary and converging
evidence for the division of participants into proficiency groups according to the C-test
scores.

6.3.2.1. Index of errors

To obtain a measure of the accuracy of the L2 production by the non-natives, all speech
samples were analysed for the number of errors/mistakes. Errors were coded in the areas
of (a) syntax, (b) case marking, (c) gender marking and (d) lexical errors (e.g. wrong
choice of word or phrase, choice of non-existing words, choice of non-German words). In
light of the syncretism of case and gender marking in German, only unambiguous errors
in case or gender marking were counted. Errors were coded by listening to the speech
samples. Subsequently, the speech samples were rechecked.

Given that the length of the individual speech samples varied considerably, an
index of errors as errors per minute was computed for the participants of each proficiency
group as defined by the C-test. Table 6.3 presents the results broken down by group. The
first row of data presents the total time of all speech samples for a given group. In the
lower rows, the numbers are errors per minute; the actual number of errors is given in
parentheses. The two rightmost columns present collapsed indices of errors for the
advanced and the near-native groups respectively; these were obtained by adding the
speech samples from each language group at the same proficiency level and then dividing
them by the total number of errors in each category.

For each language group, the near-natives produced fewer errors than the
advanced speakers in all types; for all languages collapsed, the near-natives across L1s
also produced fewer errors than the advanced speakers. However, the L1 Dutch near-
natives produced relatively more errors than their L1 English and L1 Russian near-native
counterparts for most types of errors. It is interesting to note that although the matching
proficiency groups across L1s behave similarly overall, there appears to be L1-specific
performance in some types of errors. The L1 Dutch subjects seem to make comparatively
fewer syntactic errors than their L1 English and L1 Russian counterparts, whereas the L1
Russians seem to perform better on case and gender than the two other groups. Although
the L1 groups scored comparably in the C-test, these differences in production appear to

1 From the speech samples, it cannot be assessed whether a non-target form is a mistake, i.e. a performance
inaccuracy, or an error, i.e. a systematic deviation from the target (e.g. Gass & Selinker, 2001). Since the
index is calculated with an interest to uncover systematic differences in L2 performance, I henceforth refer
to it as an index of errors, although some subset of the errors may in fact be mistakes.
reflect L1 differences in L2 German proficiency and foreshadow some of the L1-specific differences attested in the main experiments reported in this chapter.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>total time (min)</td>
<td>18:00</td>
<td>24:10</td>
<td>16:00</td>
<td>16:40</td>
<td>15:10</td>
<td>21:10</td>
<td>49:10</td>
<td>61:50</td>
</tr>
<tr>
<td>syntax errors/minute (total)</td>
<td>0.72 (13)</td>
<td>0.12 (3)</td>
<td>0.31 (5)</td>
<td>0.06 (1)</td>
<td>0.86 (13)</td>
<td>0.14 (3)</td>
<td>0.63 (31)</td>
<td>0.11 (7)</td>
</tr>
<tr>
<td>case errors/minute</td>
<td>0.39 (7)</td>
<td>0.08 (2)</td>
<td>0.56 (9)</td>
<td>0.42 (7)</td>
<td>0</td>
<td>0</td>
<td>0.33 (16)</td>
<td>0.15 (9)</td>
</tr>
<tr>
<td>gender errors/minute</td>
<td>1.33 (24)</td>
<td>0.29 (7)</td>
<td>1.44 (23)</td>
<td>0.96 (16)</td>
<td>0.40 (6)</td>
<td>0.05 (1)</td>
<td>1.08 (53)</td>
<td>0.39 (24)</td>
</tr>
<tr>
<td>lexical errors/minute</td>
<td>0.94 (17)</td>
<td>0.66 (16)</td>
<td>1.12 (18)</td>
<td>1.08 (18)</td>
<td>0.93 (14)</td>
<td>0.57 (12)</td>
<td>1.00 (49)</td>
<td>0.74 (46)</td>
</tr>
<tr>
<td>total errors/minute</td>
<td>3.39 (61)</td>
<td>1.16 (28)</td>
<td>3.43 (55)</td>
<td>2.53 (42)</td>
<td>2.19 (33)</td>
<td>0.75 (16)</td>
<td>3.04 (149)</td>
<td>1.39 (86)</td>
</tr>
</tbody>
</table>

Table 6.3. Index of errors based on speech samples. Numbers denote errors per minute, the actual number of errors is given in parentheses.

In sum, the analysis of the rate of errors in L2 production underscores that the proficiency differences obtained in the C-test also reflect accuracy in L2 production. The production data thus supply converging evidence of proficiency differences between the L2 participants.

6.3.2.2. Rating

Following the rationale in, e.g., White & Genesee (1996), all speech samples were also rated on five dimensions: (a) fluency, (b) choice of vocabulary, (c) expression, (d) mistakes and (e) accent or pronunciation. Three native speakers of German listened to all non-native speech samples, which were interspersed with speech samples of native German speakers, and rated the samples. The raters had no knowledge of the C-test score of a participant or the index of errors associated with a particular speech sample, nor did they know whether a particular sample was from a native or a non-native speaker. The raters scored each sample on the five dimensions using a ten-point rating scale. 10 was defined as the score for comprehensive native-likeness, and 0 was the score for most perceptible non-native-likeness (see Appendix C for details). The interrater reliability was calculated as an intraclass correlation across all raters and all subjects for the average scores across all dimensions, and a Cronbach’s α of 0.798 indicates a high interrater reliability. Table 6.4 lists the results for the five dimensions for each language and each proficiency group. The native speaker group scored 8.5 points or higher on each of the dimensions.
The results from the rating study show differences according to proficiency level as well as according to L1. As for proficiency, the near-native groups receive consistently higher ratings than the advanced groups. In terms of L1 differences, the L1 English and L1 Russian groups score on a par in total scores, while the L1 Dutch groups attain higher ratings. On accent, the three L1 groups pattern differently, with the L1 English groups scoring lowest, the L1 Russian groups scoring higher and the L1 Dutch groups scoring highest.

Since the rating dimensions are not discrete, and, for instance, the score for expression may partially overlap with fluency, etc., it was decided that the data would not be used for further correlational analyses within the speech ratings. In order to establish the relation between the picture-description task and the C-test, a correlational analysis was run on the z-scores of total scores of each participant in each test. There is a moderate correlation $r=0.536$ ($p<0.01$) between the scores of the C-test and the rating of the speech samples.

In sum, the C-test and the picture-description task point to substantial differences in L2 proficiency among the non-native participants. By and large, the results of the two tasks converge in yielding differences between a high-scoring ‘near-native’ group and a lower-scoring ‘advanced’ group, as defined by the C-test scores. Proficiency level will be considered as a variable in all following experiments.

### 6.4. Experiment 1: Off-line grammaticality judgements

The first experiment reported in this chapter investigates off-line knowledge of the morphosyntax of word order variation in L2 German.

The aim of this experiment is to establish whether non-natives allow for word order optionality and recognize syntactic reordering flagged by morphological, i.e. case, cues. In addition, it investigates whether the L2 groups have acquired the constraints on word order and case marking. Specifically, the research questions are:
(Q1.1) Do L2 speakers accept scrambling as a distributional option in the TL?
(Q1.2) Do L2 speakers show differences in acceptance between canonical SO and scrambled OS orders?
(Q1.3) Do L2 speakers differentiate scrambling from ungrammatical word orders, on the one hand, and from ungrammatical case marking, on the other?
(Q1.4) Are there proficiency or L1 effects?

As a reference point for studying L1 transfer effects, Table 6.5 (adapted from Table 6.1) summarizes the cross-linguistic morphosyntactic differences relevant in the context of Experiment 1.

<table>
<thead>
<tr>
<th>Target</th>
<th>L1s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GERMAN</td>
</tr>
<tr>
<td>Basic word order</td>
<td>SOV</td>
</tr>
<tr>
<td>Short scrambling</td>
<td>+</td>
</tr>
<tr>
<td>Medium scrambling</td>
<td>+</td>
</tr>
<tr>
<td>Case marking</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 6.5. Cross-linguistic differences relevant in the context of Experiment 1.

A judgement task was designed that encompassed both grammatical and ungrammatical word orders and case marking. The task was partially based on and adapted from Keller (2000b, Experiment 6). The experiment also probed off-line interface knowledge of word order optionality, in particular the interaction of syntactic reordering with discourse context and definiteness. These aspects of the experiment will be discussed in later chapters (Chapter 7). However, the complete design of the experiment will be presented here. The target pattern of responses is given in (P.1).

(P.1) Target pattern for Experiment 1
Scrambling is accepted significantly more than ungrammatical word orders and sentences ungrammatical by case marking.

6.4.1. Materials

Sentences were constructed that consisted of a matrix clause and an embedded clause containing an animate subject and an inanimate object as well as a transitive verb. A partially factorial design was employed which crossed the factors Word Order, Context and Definiteness. All factors pertained to the embedded clause. The factor Word Order had two levels: non-scrambled (SOV) (2a) and scrambled (OSV) (2b).
The Morphosyntax of Scrambling at L2 Ultimate Attainment

(2)  
   a. Maria glaubt, dass der Vater den Wagen kauft.  
      Maria thinks that the\textsubscript{NOM} father the\textsubscript{ACC} car buys
   b. Maria glaubt, dass den Wagen der Vater kauft.  
      Maria thinks that the car the father buys

The factor \textit{Context} realized effects of information structure. A \textit{wh}-question designated as focus the constituent corresponding to the \textit{wh}-element in the answer. Correspondingly, all other constituents are given information. The factor context had three levels: ALL-Focus (3a), SUBJECT-Focus (3b) and OBJECT-Focus (3c).

(3)  
   a. Was gibt's Neues?  
      ‘What's new?’
   b. Wer kauft den Wagen?  
      ‘Who buys the car?’
   c. Was kauft der Vater?  
      ‘What does the father buy?’

The factor \textit{Definiteness} pertained to the object in the embedded clause, and it had two levels: definite (4a) and indefinite (4b).

(4)  
   a. Maria glaubt, dass der Vater den Wagen kauft.  
      Maria thinks that the father the car buys
   b. Maria glaubt, dass der Vater einen Wagen kauft.  
      Maria thinks that the father a car buys

Two further factors were included: ungrammatical embedded \textit{Verb-Position Violation} (SVO/V2) crossed with \textit{Word Order} (5), and ungrammatical \textit{Case Violation} crossed with \textit{Word Order} (6). The ungrammatical word orders in (5) display correct case marking, yet the sentences instantiate illicit SVO/OVS word order in the embedded clause, since the verb is not in clause-final position. SVO/OVS coincides with V2, the mandatory word order in German matrix clauses. The ungrammatical orders in (6) display correct word order but ungrammatical case marking, i.e. nominative case on both the subject and the object. The word-order violations and the case violations were each crossed with the factor \textit{Context}.

(5)  
   a. Maria glaubt, dass der Vater kauft einen Wagen.  
      Maria thinks that the father buys a car
   b. Maria glaubt, dass den Wagen kauft der Vater.  
      Maria thinks that the car buys the father
6.4.2. Assignment to groups and lists

There were eight lexicalizations for each cell, which resulted in a total 192 stimuli. The items were distributed across lists, such that each lexicalization occurred once in each context (ALL-Focus, SUBJECT-Focus, OBJECT-Focus) per list. The lexicalizations were adapted from Keller (2000b, Experiment 6). As reported in Keller (2000b), the lexicalizations were matched for frequency on the basis of the lemmatized version of the Frankfurter Rundschau corpus.

6.4.3. Fillers

In addition to the experimental items, the task included 10 fillers, thus yielding a total of 34 items per list. The fillers were the same for each list and encompassed constructions different from those of the experimental items. The acceptability of the fillers ranged from ungrammatical word order violations to standard grammatical sentences. Like the experimental items, all fillers were embedded in wh-questions.

6.4.4. Order of presentation

The order of the experimental items and the fillers was randomized. The order of sentences was randomized automatically for each participant at the beginning of the session. The assignment of participants to one of the eight lists was determined at the beginning of each test session.

6.4.5. Procedure

The judgement task employed the method of Magnitude Estimation, a paradigm first used in psychophysics and adapted for judging linguistic stimuli by Bard, Robertson & Sorace (1996) and Cowart (1997). Magnitude Estimation aims at obtaining maximally fine-grained judgements by eliciting relative, as opposed to absolute, judgements of acceptability. Judgements are relative to a reference sentence (the modulus) and to the participant’s previous judgements. The scale of judgements is determined by the
participants; it is open-ended and has no minimum divisions. It thus allows for (potentially) infinite differentiations. The experiment was run via the Internet using WebExp software (Keller, Corley, Corley, Konieczny & Todirascu, 1998). Participants logged on to the website and completed the experiment.

The experiment proceeded in five steps: (a) display of instructions, (b) demographic questionnaire, (c) a training session involving line judgements, (d) a practice session on linguistic stimuli, and (e) the experimental session.

In the instructions, participants were first introduced to the technique of Magnitude Estimation in relation to proportional judgements of the length of lines. Participants were told to estimate the length of lines by assigning the first line, the modulus, an arbitrary number and then assign a number to each following line that expresses how long it is in proportion to the modulus. Subsequently, participants were told that this technique can be applied to acceptability judgements of sentences. Examples of sentences of various degrees of acceptability and corresponding numeric values were provided. The participants were then told that they could use any number of their choosing and that there was no upper or lower limit to the numbers they could use. Participants were encouraged to use as wide a range of numbers as they could to make distinctions of acceptability.

After starting the experiment, participants were asked to fill in a short demographic questionnaire about relevant background characteristics.

In the training session, participants were instructed on Magnitude Estimation by using judgements of physical stimuli. The participants were presented with six lines of different lengths. After seeing the first line, they were asked to assign this line (the modulus) an arbitrary number, which would serve as the reference number for the judgements of all subsequent stimuli. This line covered approximately half the width of the window in the browser. Subsequently, the participants assigned the following lines numeric values that would express the line length relative to the reference number. These lines were either shorter or longer than the modulus item, with the range of the largest to smallest being 10:1. The modulus and the corresponding numeric judgement remained in view throughout the session. Participants entered their judgements via the keyboard; upon pressing the return button, the window changed to the next item. There was no time limit for participant’s judgements.

For the practice session, six sentences were used to familiarize participants with applying Magnitude Estimation to linguistic stimuli. The sentences ranged from totally unacceptable to perfectly grammatical and were of different structures than the experimental items. A modulus item in the middle of the acceptability range was provided. As in the training session, the modulus and its corresponding numeric judgement remained on the screen throughout.

In the experimental session, participants judged 34 sentences plus the modulus. They judged the modulus first, and the modulus with its judgement remained on the
screen throughout the experimental session. The modulus was the same for all participants (see Appendix D).

6.4.6. Participants

Fifty-three non-natives of the 59 non-natives participated in this experiment. Since this experiment was run last in the series of experiments presented in this chapter and participants had to access the experiment on the Internet, unfortunately not all participants completed the task. A few subjects also encountered irresolvable problems running the experiment on their computers. For comparison, 47 native speakers of German served as a control group.

6.4.7. Data treatment

The data were normalized for each participant by dividing each of his/her judgements by his/her reference judgement. This created a common scale for all participants. As is common practice for Magnitude Estimation (Bard et al., 1996), all statistical analyses were carried out on the geometric means of the normalized judgements. A problem with comparing Magnitude Estimation data across different groups is that the ranges of judgements used by participants are not equal in each group. The conversion to geometric means takes care of this problem only to a unsatisfactory extent (e.g Featherston, 2005). Therefore, the geometric means were rescaled to make the data comparable across groups. For each group, I identified the item with the highest score and the item with the lowest score. This was typically (but not always) an uncontroversially grammatical or ungrammatical filler item. On a scale from 0 to 100, the lowest-scoring item was set equal to 0 and the highest-scoring item was set equal to 100. In this way, the ranges of judgements across groups were standardized.

6.4.8. Analysis and results

In this section, I will not consider the full set of experimental items. Since the focus of the present chapter is on the (morpho-)syntax of scrambling, I will not consider differences according to Context and Definiteness. These aspects of scrambling will be dealt with in Chapter 7. Hence, only the data for (non-)scrambled definite NPs are reported here (Table 6.6). For the present analysis, this leaves the factors Word Order, Case Violation, Verb-

---

2 The off-line judgement task was run after the processing experiments to avoid conscious reasoning about the acceptability of sentences in the judgement task to affect reading and response patterns in the on-line tasks. However, the results from the judgement task are presented first to assess the off-line knowledge of the participants.

3 The natives had been tested as part of a pilot study before, and, for logistical reasons, the native control group used in the on-line experiments could not complete the Magnitude Estimation experiment.
In order to increase the data for the remaining conditions, the first analysis establishes whether the factor Context could be collapsed, seeing that context should not have an effect on SO Word Orders, Case Violations and Verb-Position Violation. A Repeated Measures ANOVA with the factors Context (All-Focus, SUBJEC-T-Focus, OBJECT-Focus) and Sentence Type (SOV, OSV, *SVO, *OVS, *SOV, *OSV) on the acceptability judgements was run to see whether Context had an effect on judgements.

The results yield neither a significant main effect of Context (F(2,990) 1.289, p=0.276), nor an interaction of Context and Sentence Type (F(8,990) 1.164, p=0.318). Consequently, the factor Context was collapsed, so that the factor Word Order, which was crossed with the factors Case Violation and Verb-Position Violation, respectively, remained for analysis.

To establish potential group differences, a Repeated Measures ANOVA with the within-subjects factors Order (SO versus OS) and Grammaticality (grammatical, ungrammatical word order, ungrammatical case marking) was run on the geometric

<table>
<thead>
<tr>
<th></th>
<th>English Adv. (n=8)</th>
<th>English Near-Native (n=10)</th>
<th>Dutch Adv. (n=8)</th>
<th>Dutch Near-Native (n=10)</th>
<th>Russian Adv. (n=9)</th>
<th>Russian Near-Native (n=8)</th>
<th>German (n=47)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2a) SOV</td>
<td>96</td>
<td>98</td>
<td>82</td>
<td>88</td>
<td>88</td>
<td>89</td>
<td>93</td>
</tr>
<tr>
<td>(2b) OSV</td>
<td>9</td>
<td>60</td>
<td>8</td>
<td>44</td>
<td>64</td>
<td>58</td>
<td>67</td>
</tr>
<tr>
<td>(5a) *SVO/V2</td>
<td>60</td>
<td>28</td>
<td>18</td>
<td>27</td>
<td>51</td>
<td>23</td>
<td>28</td>
</tr>
<tr>
<td>(5b) *OVS/V2</td>
<td>13</td>
<td>19</td>
<td>7</td>
<td>26</td>
<td>33</td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td>(6a) *S_NOMO_NOM_V</td>
<td>64</td>
<td>14</td>
<td>65</td>
<td>8</td>
<td>23</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>(6b) *O_NOMS_NOM_V</td>
<td>10</td>
<td>13</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>10</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 6.6. Experiment 1: Acceptability ratings. Rescaling from geometric means to scale 0-100.
means of judgements. The between-subjects factors were **Language** (German, English, Dutch and Russian) and **Proficiency** (native, near-native and advanced). Since the statistical design is not fully factorial, no comparisons by item were computed for between-group comparisons.

The results of the Repeated Measures ANOVA show significant main effects of **Order** \((F(1,93) = 148.688, p<0.001)\), **Grammaticality** \((F(1,186) = 110.053, p<0.001)\) as well as an interaction of **Order** and **Grammaticality** \((F(1,186) = 3.051, p=0.026)\). Furthermore, there are significant interactions of the factor **Order** with the factors **Language** \((F(2,93) = 6.197, p=0.003)\) and **Proficiency** \((F(1,93) = 21.758, p<0.001)\). These interactions most likely indicate that the natives do not make a difference between SO and OS orders for sentences ungrammatical by case marking \((6a)\) and \((6b)\), while the non-natives, or more specifically, the advanced groups, do. Further, there is an interaction between **Grammaticality** and **Proficiency** \((F(1,93) = 6.199, p=0.004)\) as well as a three-way interaction between **Grammaticality**, **Language** and **Proficiency** \((F(2,93) = 6.124, p<0.001)\). These interactions show that the English and Dutch advanced speakers do not make a robust distinction between grammatical and ungrammatical sentences, while the Russian advanced group and all other groups do.

In light of the interactions with the factor **Proficiency** and **Language**, it was decided that separate analyses (ANOVAs and paired-samples t-tests) should be run for the natives, the near-natives and each of the advanced groups. Since up to two comparisons were made for each sentence type in the pairwise comparisons, the significance level was set according to the Bonferroni adjustment at an alpha level of .025.

**Natives**

Let us first establish the native pattern of judgements (Figure 6.1). The German natives demonstrate a main effect of **Order** \((F_{1}(1,46) = 54.453, p<0.001; F_{2}(1,7) = 42.475, p<0.001)\) and a main effect of **Grammaticality** \((F_{1}(1,46) = 127.789, p<0.001; F_{2}(1,14) = 422.041, p<0.001)\). In pairwise comparisons, the natives show a clear slope of acceptability. The canonical SOV order receives the highest acceptance, whereas the scrambled OSV order ranks significantly lower in acceptability; a pairwise comparison between SOV and OSV orders elicits a significant difference \((F_{1}(1,46) = 9.516, p<0.001; F_{2}(1,7) = 9.334, p<0.001)\).
Yet, the scrambled OSV order is judged to be significantly more acceptable than the ungrammatical V2 orders ($F_1(1,46) = 7.710, p<0.001$; $F_2(1,7) = 14.270, p<0.001$). The natives further make a distinction in their acceptability judgements among ungrammatical sentences: The ungrammatical SVO order is judged to be more acceptable than the SOV order showing a case violation ($F_1(1,46) = 5.435, p<0.001$; $F_2(1,7) = 6.525, p<0.001$). However, natives do not differentiate between incorrectly case-marked sentences according to word order: neither is the difference between SOV and OSV orders with nominative-marked NPs significant ($F_1(1,46) = 2.077, p=0.043$; $F_2(1,7) = 2.348, p=0.051$), nor the difference between SVO/V2 and OVS/V2 ($F_1(1,46) = -1.410, p=0.165$; $F_2(1,7) = -1.330, p=0.225$). Arguably, the higher acceptability of V2 in embedded clauses compared to case violations reflects the fact that V2 is the mandatory order in main clauses. In addition, some complementizers in German also require (e.g. *denn, ‘as’) or allow (e.g. *weil, ‘because’) V2 in embedded clauses, and complementizerless embedded clauses require V2 (e.g. Grewendorf, 1988).

In sum, the natives differentiate between the preferred SOV order and the less preferred OSV order among grammatical sentence types; the less preferred OSV order, however, is judged to be significantly more acceptable than ungrammatical word orders or sentences ungrammatical by case marking.

**Near-Natives**

Like the natives, the near-natives show an effect of *Order* ($F(1,25) = 23.426, p<0.001$) as well as an effect of *Grammaticality* ($F(2,25) = 32.563, p<0.001$). There is no interaction with the factor *Language* of either the factor *Order* ($F(2,25) = 0.041, p=0.505$) or the factors *Order* and *Grammaticality* ($F(4,25) = 1.475, p=0.224$). Hence, the L1 near-native groups are collapsed for analysis. Figure 6.2 shows the results for all near-natives.
In subsequent pairwise comparisons, the near-natives demonstrate an acceptability slope that is very similar to that of the natives. The canonical SOV order is rated best and the scrambled OSV receives lower acceptability ratings; a pairwise comparison between SOV and OSV orders elicits a significant difference ($F_1(1,27) = 5.722$, $p<0.001$; $F_2(1,7) = 9.730$, $p<0.001$). In turn, the scrambled OSV order is judged to be significantly more acceptable than the ungrammatical V2 orders ($F_1(1,27) = 3.108$, $p=0.004$; $F_2(1,7) = 4.628$, $p=0.002$). Among the ungrammatical sentences, the near-natives prefer SVO/V2 over SOV orders ungrammatical by case marking ($F_1(1,27) = 2.351$, $p=0.026$; $F_2(1,7) = 0.931$, $p=0.383$). However, the near-natives do not differentiate between incorrectly case-marked sentences according to word order: the difference between SOV and OSV with nominative-marked NPs is not significant ($F_1(1,27) = 1.781$, $p=0.086$; $F_2(1,7) = 1.305$, $p=0.233$). The difference between SVO/V2 and OVS/V2 is also not significant ($F_1(1,27) = 1.037$, $p=0.309$; $F_2(1,7) = 0.608$, $p=0.549$).

Summarizing, the near-natives perform in a target-like manner: They rate the scrambled OSV order lower than the canonical SOV; yet, they judge the less preferred OSV order to be significantly more acceptable than ungrammatical word orders or sentences ungrammatical by case marking.

Advanced Speakers

In the between-subjects ANOVA for the advanced group, there is a main effect of Order ($F_1(1,22) = 52.152$, $p<0.001$) as well as a main effect of Grammaticality ($F_1(2,22) = 16.031$, $p<0.001$). Crucially, there is also an interaction of Grammaticality with the factor Language ($F_1(2,22) = 6.611$, $p=0.002$). This interaction shows that the L1 English and L1 Dutch advanced groups show divergent judgement patterns compared to the L1 Russian advanced group. As a consequence, separate analyses were run for the advanced groups by language.
**L1 English advanced group**

The L1 English advanced group shows a main effect of *Order* \( (F_1(17) 12.183, p=0.010; F_2(1,7) 87.365, p<0.001) \), yet no main effect of *Grammaticality* \( (F_1(1,7) 1.319, p=0.291; F_2(2,14) 0.897, p=0.395) \).

![Graph showing the results of the magnitude estimation task for L1 English advanced group.](image)

**Figure 6.3.** Experiment 1: Results (degree of acceptability) of Magnitude Estimation task for L1 English advanced group \( (n=8) \).

Pairwise comparisons bear out that they do not distinguish between the grammatical OSV order \( (2b) \) and sentences that are ungrammatical by word order \( (5b) \) \( (F_1(17) -0.753, p=0.476; F_2(1,7) -0.462, p=0.658) \) or by case marking \( (6b) \) \( (F_1(1,7) -0.161, p=0.876; F_2(1,7) -0.076, p=0.941) \). Moreover, they overaccept the ungrammatical SVO order; in fact, there is no difference between the high acceptance of SVO orders and ungrammatical case marking in the SOV order \( (F_1(1,7) -0.145, p=0.889; F_2(1,7) 0.088, p=0.932) \).

**L1 Dutch advanced group**

The L1 Dutch advanced group shows a main effect of *Order* \( (F_1(1,7) 62.802, p<0.001; F_2(1,7) 47.430, p=0.001) \) and a main effect of *Grammaticality* in the analysis by subjects \( (F_1(1,7) 7.790, p=0.011; F_2(2,14) 3.353, p=0.078) \). As their L1 English counterparts, they do not make a distinction between grammatical OSV orders and ungrammatical sentences by word order \( (5b) \) \( (F_1(1,7) 0.146, p=0.888; F_2(1,7) 0.638, p=0.544) \) or by case marking \( (6b) \) \( (F_1(1,7) 0.654, p=0.534; F_2(1,7) 0.397, p=0.703) \); yet, unlike the English group, they do not equally overaccept SVO orders. These are judged marginally significantly differently compared to the case violations \( (F_1(1,7) -2.648, p=0.033; F_2(1,7) -2.148, p=0.075) \).
Figure 6.4. Experiment 1: Results (degree of acceptability) of Magnitude Estimation task for L1 Dutch advanced group (n=8).

L1 Russian advanced group

The L1 Russian advanced speakers demonstrate main effects of *Order* ($F_1(1,8) = 27.490$, $p=0.001$; $F_2(1,7) = 7.396$, $p=0.030$) as well as *Grammaticality* ($F_1(1,8) = 15.753$, $p=0.001$; $F_2(2,14) = 13.114$, $p=0.003$).

Figure 6.5. Experiment 1: Results (degree of acceptability) of Magnitude Estimation task for L1 Russian advanced group (n=9).

In subsequent pairwise comparisons, the L1 Russian advanced speakers make a distinction between the canonical SOV and the scrambled OSV order ($F_1(1,8) = 2.322$, $p=0.049$; $F_2(1,7) = 3.680$, $p=0.008$) as well as between the scrambled OSV order and ungrammatical V2 orders in the analysis by subjects ($F_1(1,8) = 2.795$, $p=0.023$; $F_2(1,7) = 1.375$, $p=0.211$). They also make target-like distinctions between the ungrammatical sentence types, even though Figure 6.3 illustrates that SO orders are rated higher across sentence types.
In sum, all near-natives manifest target-like relative acceptability judgements on scrambling and case-marking, while the advanced L2 groups differ according to L1: The L1 Russians demonstrate target-like judgements, while the L1 English and L1 Dutch advanced groups deviate from target judgement patterns.

6.4.9. Discussion

Experiment 1 yielded the following results:

- The native group shows a clear slope of acceptability with the canonical SOV order being the most acceptable and the scrambled OSV order being less preferred. However, the grammatical OSV order is rated significantly more acceptable than ungrammatical word orders or sentences ungrammatical by case marking.
- The near-native groups show target judgement patterns. Moreover, the near-natives demonstrate analogous acceptability patterns across L1s.
- The advanced groups show different judgement patterns according to L1:
  - The L1 Russian group patterns along with the near-natives and the natives.
  - The L1 English advanced group accepts all subject-first orders, irrespective of whether they are grammatical or ungrammatical by word order or by case marking. In turn, the L1 English advanced group does not accept object-initial orders, again irrespective of whether they are grammatical or ungrammatical by word order or by case marking.
  - The L1 Dutch advanced group does not accept object-initial orders, irrespective of their (un-)grammaticality. For subject-initial sentences, the Dutch advanced group accepts the canonical SOV order, yet also the SOV order with ungrammatical case marking. However, the L1 Dutch advanced group shows a target-like low acceptance of SVO in embedded clauses.

Despite high levels of general proficiency, the L1 English and L1 Dutch advanced groups have not converged on the target-language grammar in the area of word order optionality. This contrasts somewhat with the findings of previous research that mid- to high-proficiency L2 learners recognize the distributional options of scrambling (e.g. Iwasaki, 2003; Jackson, 2005; Koda, 1993; Neeleman & Weerman, 1997). The non-convergence in the present experiment might partly be due to the Magnitude Estimation judgement task that requires subjects to rate the relative acceptability of sentences rather than declaring them either grammatical or ungrammatical. Making relative judgements poses greater difficulty for the participants, so that the Magnitude Estimation task is sensitive to potential nuances in judgements that cannot be elicited in, e.g., binary judgement tasks.
L1 effects

L1 effects surface at advanced proficiency levels. These are arguably due to L1 differences in case marking on full NPs: Russian uses case marking for syntactic function assignment. As in their L1s, the L1 English and L1 Dutch groups do not use case marking for establishing the syntactic function of NPs. Instead, their blanket preference for subject-initial orders, irrespective of case marking, suggests that the L1 English and Dutch groups prioritize animacy information or linear order for indicating syntactic relations.

Unfortunately, it is not possible to disentangle the role of animacy and linear order on the basis of the present data. In order to yield interpretable results, the sentences of the judgement task contained animate subjects and inanimate objects; using nouns matched on animacy would not have permitted us to ascertain whether participants make a difference between non-scrambled and scrambled orders in judgements. Hence, it is not clear whether the L1 English and L1 Dutch advanced groups rule out an inanimate (object) NP in first argument position because it cannot be the subject due to the nonsensical reading this would give or whether they rule them out solely because of inanimacy.

Other studies, however, furnish evidence that lower-level L2 speakers rely on a subject-initial strategy, rather than on animacy information. In off-line judgements on German scrambling elicited by mid-proficient L1 English learners, Jackson (2005) finds that the L2 participants are not sensitive to animacy differences; rather, they demonstrate a blanket preference for SO orders over OS orders, even if the subject was inanimate and the object animate. Experiments on L2 Spanish word order options by VanPatten (1996) show that linear order, not animacy, accounts for the subject-first preference of L1 English learners. Similarly, studies on cross-linguistic transfer of interpretive cues within the Competition Model (MacWhinney & Bates, 1989), summarized in MacWhinney (2005), attest that English-Dutch and Dutch-English L2 learners prioritize word order, yet not animacy, in the L2 interpretation of syntactic functions. In a similar vein, Kempe & MacWhinney (1998) argue that L1 English learners of German use word order cues more than animacy cues for detecting the syntactic function of NPs. It thus seems reasonable to conclude that the advanced learners are guided by a subject-initial strategy in their judgements, rather than by an animacy-based strategy.

For the advanced L1 English and L1 Dutch groups, the subject-first strategy overrules conflicting case-marking information. Such a lack of sensitivity to case marking even in off-line tasks echoes the results of grammaticality judgement studies on mid- to high-proficiency L1 English learners of case-marking languages like Spanish (Montrul, 1999), Turkish (Gürel, 2000) and German (Hopp & Schwartz, 2002).

In the present experiment, it is unlikely that the failure of the L1 English and L1 Dutch advanced learners to employ case information stems from lack of knowledge of
inflectional paradigms or from their inability to identify the form-to-function mappings of the particular case markers used in this experiment. The case markers used represent the prototypical nominative (der) and accusative (den) markers in German and are subject to extensive instruction. Indeed, in the production data of the participants as reflected in the error index (Section 4.2.2.1), it is found that (unambiguous) case errors are the least frequent type of error made. Rather, it appears that word order, perhaps in conjunction with animacy information, overrides the low-salient case information even in off-line judgements in that the advanced L1 English and L1 Dutch speakers employ L1-based strategies in judging L2 German sentences.

Further effects of L1 transfer are attested in the asymmetric judgements of the L1 English group versus the L1 Dutch group on ungrammatical embedded SVO sentences. Whereas the advanced L1 English group accepts embedded SVO by transferring the structure of English to German, the advanced L1 Dutch group rules out embedded SVO in line with the analogous SOV structure of Dutch embedded clauses.

Proficiency effects

At the near-native level, increases in proficiency cancel all of these L1 effects in that all three near-native groups converge on target acceptability slopes in judgements. At the high end of proficiency, L2 learners of various L1 backgrounds thus come to acquire knowledge of comparatively infrequent syntactic constructions in German and robustly differentiate them from ungrammatical word orders and ungrammatical case marking. In terms of the syntactic analysis of scrambling developed in Chapter 3, the results of Experiment 1 show that L2 speakers can come to acquire (L1 English) or reconfigure (L1 Dutch) an uninterpretable scrambling feature to accommodate pre-subject scrambling in L2 German.

In sum, the off-line task hence furnishes evidence that convergence on the morphosyntactic properties of German word order optionality is attainable in grammatical knowledge for adult L2 learners of various L1 backgrounds, irrespective of whether the L1 shows overt case marking (Russian) or not (English, Dutch) and irrespective of whether the L1 is underlyingly SVO (English, Russian) or SOV (Dutch). The remaining experiments in this chapter examine whether convergence on the morphosyntax of scrambling in off-line judgements extends to on-line processing. In particular, it will be investigated whether the near-native groups who show target-like knowledge of case marking and scrambling off-line also use case marking robustly for syntactic reanalysis in L2 processing.
6.5. Processing experiments: Recap

The overarching question for the processing experiments is whether L2 speakers demonstrate target-like reanalysis in the processing of the morphosyntax of scrambling. As discussed in Chapter 4, the native processing of subject-object ambiguities in German shows a robust pattern of reanalysis reflecting the interaction of phrase-structure-based parsing principles and morphosyntactic features, summarized in (7):

(7) Native processing of the morphosyntax of subject-object ambiguities

(a) Native processing is guided by a structural subject-first preference that reflects the operation of economy-based parsing principles in incremental phrase-structure building (e.g. the Minimal Chain Principle, DeVincenzi, 1991).
(b) OS orders invoke syntactic reanalysis. Reanalysis is measurable in global as well as local reading delays at the points of morphological disambiguation to the OS order.
(c) In grammatical sentences, OS orders disambiguated by number marking on the verb evince greater processing cost than OS orders disambiguated by case marking, while the inverse holds in ungrammatical sentences (see Table 6.7). This contrast reflects the different grammatical informativity of morphosyntactic features for syntactic reanalysis (Fodor & Inoue, 2000; Meng & Bader, 2000).

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Verbal Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical OS sentences:</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>Ungrammatical sentences:</td>
<td>high</td>
<td>low</td>
</tr>
</tbody>
</table>

Table 6.7. Interaction of feature of disambiguation and sentence type in parsing.

In the remainder of this chapter, two processing experiments will be reported that test for the native target pattern of processing in (7):

Experiment 2: A self-paced reading task on scrambling. This task probes the extent to which non-natives evince incremental effects of reanalysis, i.e. immediately use morphological information for phrase-structure revision.

Experiment 3: A speeded-grammaticality judgement task. This task investigates the extent to which non-natives show target-like interactions of processing difficulty and morphosyntactic feature type (i.e. case versus subject-verb agreement).
In conjunction, these experiments can furnish evidence on the processing of morphosyntax in advanced to near-native L2 acquisition.

6.6. Experiment 2: Self-Paced Reading

In a self-paced reading task, subjects read sentences segment by segment at their own pace. The rationale of this paradigm is that increased processing effort, i.e. reanalysis effects, can be detected locally in higher reading times on a given segment compared to the same segment in a control condition. Self-paced reading can thus show incremental effects of processing on top of global effects at the sentence level.

The cross-linguistic differences in scrambling that are relevant in the context of the present experiment are listed in Table 6.8.

<table>
<thead>
<tr>
<th></th>
<th>Target</th>
<th>L1s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GERMAN</td>
<td>ENGLISH</td>
</tr>
<tr>
<td>Scrambling</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Overt case marking</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Overt verbal agreement</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 6.8. Cross-linguistic differences relevant in the context of Experiment 2.

Note that there are cross-linguistic differences in the availability of scrambling and morphological case marking in the four languages. All of them instantiate morphologically marked subject-verb agreement. Against the background of the cross-linguistic differences, the research questions for Experiment 2 are:

(Q2.1) Do L2ers show incremental reanalysis effects to the OS order on the regions of disambiguation?
(Q2.2) Do L2ers show differences in processing difficulty depending on the type of morphological disambiguation (case vs verbal agreement)?
(Q2.3) Are there L1 differences in reanalysis depending on whether the L1 instantiates scrambling (English versus Dutch and Russian) or depending on whether the L1 instantiates scrambling and case marking (English and Dutch versus Russian)?
(Q2.4.) Are there proficiency differences in the processing of scrambling constructions?

The target pattern for Experiment 2 is given in (P.2).
Chapter 6

Target pattern for Experiment 2

a. Incremental reanalysis effects surface in reading slowdowns on the points of morphological disambiguation to the OS order, i.e. on the first accusative-marked NP for sentences disambiguated by case and on the finite verb for sentences disambiguated by verbal agreement.

b. Sentences disambiguated to an OS order by case marking receive higher comprehension accuracy than OS sentences disambiguated by verbal agreement (see Table 6.7).

6.6.1. Materials

Twenty-four quadruplets of sentences were constructed. Half of the quadruplets were disambiguated by case on determiners (e.g. (8)) and half by number marking on the verb (e.g. (9)) (Factor Ambiguity). Sentences within each quadruple set were in SO or OS order (Factor Order). Two additional versions within each quadruple set were constructed by reversing the position of the nouns (i.e. N1-N2 and N2-N1), so that any potential effect of lexical semantics or pragmatics of the SO and OS manipulation would be completely matched. The reversal of noun position was not treated as a factor in the statistical design. All sentences were initiated by a matrix clause. Sentences disambiguated by number marking contained two NPs, one plural and one singular, both with ambiguous case inflection, and the finite verb was plural in half of the sentences and singular in the other. Sentences disambiguated by case contained two NPs, one unambiguously marked for nominative, the other unambiguously marked for accusative. Examples of two quadruplet of the experimental items are given in (8) and (9).

(8) a. Er denkt, dass der Physiker am Freitag den Chemiker gegrüsst hat. SO He thinks that theNOM physicist on Friday theACC chemist greeted has
b. Er denkt, dass den Physiker am Freitag der Chemiker gegrüsst hat. OS
c. Er denkt, dass der Chemiker am Freitag den Physiker gegrüsst hat. SO
d. Er denkt, dass den Chemiker am Freitag der Physiker gegrüsst hat. OS

(9) a. Sie sagt, dass die Baronin am Freitag die Bankiers eingeladen hat. SO She says that the baronessSG on Friday thebankersPL invited has
b. Sie sagt, dass die Baronin am Freitag die Bankiers eingeladen haben. OS She says that the baronessSG on Friday thebankersPL invited have
c. Sie sagt, dass die Bankiers am Freitag die Baronin eingeladen haben. SO
d. Sie sagt, dass die Bankiers am Freitag die Baronin eingeladen hat. OS
All embedded sentences were in the present perfect tense. For sentences disambiguated by number marking on the verb, the auxiliary hat (‘has’) or haben (‘have’) would disambiguate the sentences. I opted for the perfect tense to avoid variation in verb inflection on the critical segment. For this reason, only active verbs that form the perfect tenses with haben (‘have’) were used in the embedded clause. All verbs were agentive verbs or accusative-marking psych verbs, for which native processing patterns are identical (Scheepers et al., 2000). In order to ensure that the NPs within and across conditions would be of similar length, NPs that were between 7 and 9 characters long were chosen; the verbs were also between 7 and 9 characters in length. The NPs were further matched for gender and animacy. In the items for (8), the NPs were also matched for number.

6.6.2. Comprehension Questions

In order to ensure that participants read the sentences properly and in order to check their comprehension of the sentences, each item was followed by a comprehension statement. For the experimental items, this statement was a main clause in SO order that contained both NPs and the verb of the preceding experimental item. An example of a comprehension sentence in response to (8a) or (8b) is given in (10).

(10) Der Physiker hat den Chemiker gegrüssst.
The physicist has the chemist greeted

For half of the experimental items, the comprehension statement was a correct rendering of the experimental item; for the other half, the comprehension sentence was an incorrect rendering of the experimental item.

6.6.3. Plausibility and reversibility

The items were constructed with attention to the plausibility and the semantico-pragmatic reversibility of the predicates: A sentence with NP1 for the subject and NP2 for the object was to be as plausible as the sentence with NP2 for the subject and NP1 for the object. The reversibility of sentences was tested in an off-line rating study. In a 100-item rating task, 54 potential experimental items were tested. The items were presented as embedded clauses initiated by a matrix clause, such as in (8/9a). The sentences were always in SO order. The items were interspersed with fillers and potential items of a different structure for other experiments. Two subject lists were created by reversing the order of NP1 and NP2, i.e. NP1-NP2 and NP2-NP1. The sentences were judged by 16 participants, 8 for each list. The participants were all native speakers of German. Participants were asked to rate the plausibility of the sentences on a scale ranging from 1 to 5, with 1 being defined
as ‘not plausible’ and 5 as ‘plausible’. The experiment yielded results as to the plausibility and the reversibility of items. On the basis of these ratings, the 24 most plausible and reversible items were selected for use in the self-paced reading task. As for plausibility, only items that gathered mean plausibility ratings above 3 were used; as for reversibility, only items that did not differ between NP1-NP2 and NP2-NP1 combinations by more than 1 between the group means were included in the experiment. The mean plausibility score for the NP1-NP2 items in List 1 was 4.2 and 4.0 for the NP2-NP1 items in List 2. The difference between the two lists is not significant (p>0.05). The complete set of experimental materials for Experiment 2 is given in Appendix E.

6.6.4. Assignment to groups and lists

The 24 quadruplets were divided into four item groups of six sets each. On the basis of these four item groups, four subject lists were created, using a Latin Square design. Sentences in item group 1 appeared in the subject-object N1-N2 version in list 1, in the subject-object version N2-N1 in list 2, in the object-subject N1-N2 version in list 3 and in the object-subject N2-N1 version in list 4. Sentences in group 2 appeared in the subject-object N2-N1 version in list 1, in the object-subject N1-N2 version in list 2 and so on. In this way, each list contained an equal number of sentences in each condition, and no list contained more than one version of each sentence. In order to ensure that all relatively infrequent, implausible or irreversible items (see above) would not appear in one condition on a list, the item groups were matched on (a) mean plausibility and reversibility scores and (b) logarithmic frequency of NPs and the verb (CELEX database; Baayen, Piepenbrock & Gulikers, 1995). Participants were pseudorandomly assigned to lists such that each list was read by a similar number of participants.

6.6.5. Fillers

On top of the 24 experimental items, the task included 78 fillers, thus yielding a total of 102 items. The fillers were the same for each list and encompassed constructions different from those of the experimental items. Each filler was followed by a comprehension statement; for half the fillers, the comprehension statement was correct, for the other half, it was incorrect.

6.6.6. Order of presentation

The order of the experimental items and the fillers was randomized. The order of sentences was randomized automatically for each participant at the beginning of the task.
6.6.7. Procedure

The paradigm was a non-cumulative Moving Windows task (Just et al., 1982) and was run using E-Prime software (Schneider, Eschman & Zuccolotto, 2002). Participants were tested individually. They sat in front of a laptop computer with a 15-inch TFT screen. The keyboard was covered with a blind that left only the spacebar and two keys visible. The spacebar was the ‘Go’ key. The two other keys were slightly offset from the rest of the keyboard on the bottom right. The left key was marked with a green sticker (the ‘yes’ key) and the right key was marked with a red sticker (the ‘no’ key).

A trial was initiated by pressing the ‘Go’ key. Each trial sentence was preceded by a fixation star in the centre of the screen. When the subjects pressed the ‘Go’ key, the first segment appeared at the left edge in the vertical middle of the screen. At the next push of the key, the first segment disappeared and the second segment appeared to the right of the now-gone segment. All sentences fit on one line. The presentation of the sentences included punctuation. Spelling and punctuation were according to the reformed spelling conventions (Duden, 1996). The end of a sentence was signaled by a full stop following the last word in the final segment. The experimental sentences were divided into seven segments (11).

(11=8a) Matrix COMP NP1 adverbial NP2 V-part V-fin
Er denkt | dass | der Physiker | am Freitag | den Chemiker | gegrüsst | hat.

All text was presented in Courier New Font, font size 14, in white letters against a black background. Once the last word of the final segment had been read, the comprehension statement, e.g. (10), was presented in its entirety. The comprehension statement was presented in Courier New Font, font size 14, in yellow letters against a black background in the centre of the screen. At this point, the participant had to decide whether the comprehension statement matched the content of the experimental sentence by pressing the ‘yes’ key or the ‘no’ key. Once the response had been given, the screen changed to display the fixation star, and the following trial could be initiated by pressing the space bar.

The participants were told that the experiment was about L2 reading comprehension and were instructed to read for comprehension. They were asked to read at their normal reading speed. The presentation of the actual items was preceded by a page of instructions and six practice items. Feedback to the participants’ responses in the practice items was given by the experimenter. Once the participants had understood the procedure and could operate the buttons, the actual experiment was started. The participants did not receive any feedback during the experimental session. Most participants completed the task within 14 minutes.
6.6.8. Experiment 2: Analysis and data removal

For each sentence type, reading times faster than 100ms or slower than 5000ms were treated as missing data. In addition, reading times of a participant for a segment above or below two standard deviations of the group mean for that segment were trimmed to the group mean of the segment plus or minus two standard deviations, respectively. This affected less than 2% of the trials in each group. The analyses of reading times were run both for raw reading times and for residual reading times.

Raw reading times show some variability depending on differences in reading speed among individuals and differences in the length of segments. In order to filter out these effects that introduce noise into the data, residual reading times were calculated (e.g. Ferreira & Clifton, 1986; Kaan, 1997). Residual reading times were calculated in the following way. On the basis of the reading times of all segments of all items, including the filler items, a linear regression was estimated for each participant with Reading Time as the dependent variable and Length of Segment in number of characters as the explanatory variable. For each participant, residual reading times were then obtained by subtracting the reading times estimated by the individual linear regression from the actual reading times. Hence, if a segment shows a positive number as the residual reading time, its reading time is slower than predicted on the basis of the length of the word; if a segment shows a negative number, its reading time is faster than predicted.

A mixed three-way Repeated Measures ANOVA with Order and Ambiguity as within-subjects factors and with Language (German, English, Dutch, Russian) and Proficiency (native, advanced L2 and near-native L2) as between-subjects factors was performed for each segment separately on both raw and residual reading times. Finding interactions with the factor Language would indicate that (some) non-native processing is different from native processing or that there are L1 differences among the L2 groups. Finding interactions with the factor Proficiency would indicate that processing differs according to proficiency level. Post-hoc two-way analyses were computed to investigate the causes of interactions of the within-subject factors and Language or Proficiency. Results are reported as significant if the p-value is less than .05, although effects with a p-value lower than .1 are reported if they are of interest. For the post-hoc analyses, the significance level was adjusted using Bonferroni adjusted alpha levels of .025 per test.

For the reading times, analyses were run on all items and on only those items for which participants gave correct comprehension responses. There were no major differences between these analyses. The few differences are indicated in Table 6.11 below and will be reported in the text if relevant for the discussion. In the following, analyses are reported for all items regardless of comprehension accuracy, since at issue is how (non-native) participants typically read and process sentences, rather than how they read and process the proportion of sentences for which they correctly answer comprehension questions. Considering only this latter subset of sentences might give a
picture of L2 performance which is in fact characteristic of only a small percentage of items. Moreover, for the question as to whether non-natives use syntactic information in on-line reading of the TL, it is irrelevant if participants correctly respond to a comprehension question that they read and answer subsequent to reading the item sentence. Answering questions about linguistic stimuli involves additional processes, preferences and strategies compared to the process of reading sentences. Hence, errors in the comprehension questions may be caused by many factors and, for this reason, the participants’ accuracy scores for answering the comprehension questions will be analysed and discussed separately (see also Kaan, 1997).

In addition, the accuracy of answering the comprehension questions was not distributed equally across conditions, such that excluding items with incorrectly answered comprehension questions would violate the assumption of homogeneity of variance. This would increase the likelihood of a Type I error, i.e. rejecting the null hypothesis when in fact it is true. Moreover, looking at only the correctly answered items would lead to empty cells in both the subject and item analyses in particular for the non-native groups, since, for some sentence types, some participants answered all questions incorrectly, and, for some items, all questions were answered incorrectly.

Finally, note that for the sentences disambiguated by verbal agreement, the region of disambiguation coincides with the sentence-final segment. Psycholinguistic research reports that additional processing effort obtains in sentence-final position, so called wrap-up effects. The nature of wrap-up effects is ill-understood (Balogh, Zurif, Prather, Swinney & Finkel, 1998; Rayner, Kambe & Duffy, 2000); they are usually taken to involve general information integration processes to construe an interpretation of a sentence or to check its well-formedness. Regardless of their function, wrap-up effects can interfere with syntactic processing components, such that the critical regions are usually placed before the sentence-final position. In the present experiments, this would have necessitated adding an additional clause to the embedded clause, which would have rendered the sentences overly long and unnatural. In order to tease apart specific local slowdowns for OS orders disambiguated on the sentence-final segment from general wrap-up effects, I will consider in the statistical analysis whether there is an interaction of Order and Ambiguity on the relevant segment. In this way, effects locally delimited to OS orders disambiguated on the sentence-final segment can be distinguished from (a) general wrap-up effects or (b) wrap-up or end-of-sentence effects specific to OS orders. Both types of wrap-up effects would give rise to main effects.

6.6.9. Participants

All 59 non-native participants and 20 natives took part in Experiments 2. One participant in the near-native L1 English group was removed from analysis because the data set was incomplete.
6.6.10. Results

6.6.10.1. Comprehension questions

On top of the reading times of sentences, the comprehension accuracy of sentences gives insight into global processing behaviour: Typically, a sentence that elicits processing difficulty in reading leads to both lower accuracy and slower response times in the comprehension questions. Table 6.9 lists the accuracy scores for each sentence type by group. The high accuracy rates for SO orders show that subjects read attentively and completed the task properly. By comparison to SO orders, scrambled OS orders lead to lower comprehension accuracies. OS orders thus give rise to stronger comprehension difficulty after disambiguation than their SO counterparts.

<table>
<thead>
<tr>
<th></th>
<th>ENGLISH Advanced (n=10)</th>
<th>ENGLISH Near-Native (n=9)</th>
<th>DUTCH Advanced (n=10)</th>
<th>DUTCH Near-Native (n=10)</th>
<th>RUSSIAN Advanced (n=10)</th>
<th>RUSSIAN Near-Native (n=9)</th>
<th>GERMAN (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (8a): SO Case</td>
<td>88% (53/60)</td>
<td>87% (47/54)</td>
<td>95% (57/60)</td>
<td>90% (54/60)</td>
<td>85% (51/60)</td>
<td>91% (49/54)</td>
<td>86% (103/120)</td>
</tr>
<tr>
<td>Accuracy (8b): OS Case</td>
<td>42% (25/60)</td>
<td>74% (40/54)</td>
<td>22% (13/60)</td>
<td>60% (36/60)</td>
<td>42% (25/60)</td>
<td>61% (33/54)</td>
<td>56% (67/120)</td>
</tr>
<tr>
<td>Accuracy (9a): SO Verbal Agr.</td>
<td>95% (57/60)</td>
<td>91% (49/54)</td>
<td>95% (57/60)</td>
<td>60% (56/60)</td>
<td>92% (55/60)</td>
<td>96% (52/54)</td>
<td>93% (111/120)</td>
</tr>
<tr>
<td>Accuracy (9b): OS Verbal Agr.</td>
<td>23% (14/60)</td>
<td>44% (24/54)</td>
<td>28% (17/60)</td>
<td>45% (27/60)</td>
<td>23% (14/60)</td>
<td>41% (22/54)</td>
<td>28% (34/120)</td>
</tr>
</tbody>
</table>

Table 6.9. Experiment 2: Comprehension accuracy in percent. Numbers in parentheses are absolute responses.
The between-subjects ANOVA reveals a significant main effect of Order (F(1,71) 293.988, p<0.001), a main effect of Ambiguity (F(1,71) 18.020, p<0.001) and an interaction of Order and Ambiguity (F(1,71) 52.654, p<0.001). The interaction between Order and Ambiguity reflects the proportionally lower accuracy on OS orders disambiguated by verbal agreement compared to OS orders disambiguated by case marking. There was also an interaction of Order and Proficiency (F(1,71) 13.800, p<0.001). This interaction appears to reflect the lower comprehension accuracy on OS orders by the advanced groups vis à vis the (near-)natives. In addition, a three-way interaction between Order, Ambiguity and Language reaches significance (F(2,71) 3.729, p=0.029). This three-way interaction most likely arises due to the L1 Dutch group not making a distinction in accuracy between sentences disambiguated by case and sentences disambiguated by verbal agreement.

6.6.10.2. Response times to comprehension questions

The response times were analysed for items that were answered correctly. Typically, sentences that prove hard to understand also elicit longer response times, such that the analysis of response times can add to the analysis of comprehension accuracy. Since response times were limited to maximally 4000ms, participants had to make quick decisions. The speed at which they arrived at the target comprehension allows for an initial insight into processing difficulty. A Repeated Measures ANOVA with the same factors as for response accuracy was carried out. Since some participants did not correctly answer any questions for some sentence types, there were 59 participants whose data were analysed. Table 6.10 shows the reaction times for each sentence type by group.

The between-groups ANOVA yields a main effect of Order (F(1,52) 17.756, p<0.001). There is also an interaction of Order and Language (F(2,52) 3.516, p=0.037), which arguably reflects the fact that for disambiguation by case, the L1 English and L1 Russian groups do not show significantly slower response times for OS orders than for SO orders. No further effects reach significance. In sum, OS orders are lower in comprehension accuracy and take longer to judge than SO orders. In terms of global processing measures of sentence comprehension, OS orders require increased processing effort and lead to decreased global processing accuracy compared to SO orders.
Table 6.10. Experiment 2: Response Times (in milliseconds) to comprehension question.

6.6.10.3. Reading times

In order to explore incremental effects of processing difficulty, the reading times were analysed for each segment in the self-paced reading task. For the reading times, Table 6.11 charts the significant effects and notes differences between the analyses for all items and the analyses for only those items for which comprehension questions were answered correctly. Reading times of each individual group are listed in Appendix F.

The total reading times show a significant effect of Order (F(1,71) 29.849, p<0.001), which reflects the fact that OS sentences evince longer reading times overall than their SO counterparts.
As for individual segments, there were no significant effects on segments 1 or 2, i.e. the lead-in main clause. On segment 3 (NP1), there are main effects of Order (F(1,71) 23.604, p<0.001) and Ambiguity (F(1,71) 4.951, p=0.010) as well as an interaction of Order and Ambiguity (F(1,71) 28.024, p<0.001). This interaction demonstrates that, in segment 3, OS orders disambiguated by case evince significantly longer reading times compared to the other sentence types. In addition, there is an interaction of Ambiguity and Language (F(2,71) 4.951, p=0.010) that arguably reflects the slower reading times of case-ambiguous NPs by the non-native groups compared to the natives. On segment 4 (ADV), there is a significant main effect of Ambiguity in the analysis of the residual reading times (F(1,71) 28.899, p<0.001). In the analysis of the raw reading times of segment 4, there are interactions of Ambiguity and Language (F(2,71) 6.548, p=0.002) as well as a three-way interaction of Order, Ambiguity and Language (F(2,71) 6.116, p=0.004). These effects seem to spill over from segment 3. In addition, the interaction between Order and Proficiency reaches marginal significance in the analysis by subjects (F(1,71) 2.855, p=0.095; F(1,161) 1.269, p=0.262). Arguably, this is an effect spilling over from segment 3, where the interaction of Order and Proficiency just fails to reach marginal significance (F(1,71) 2.608, p=0.111; F(1,161) 2.208, p=0.139).

In light of the spillover effects from segment 3 onto segment 4, the reading times were collapsed for all subsequent analyses. For segment 3 and 4 collapsed, henceforth called Segment 3&4 (NP1 & ADV), there are main effects of Order (F(1,71) 17.452, p<0.001) and Ambiguity (F(1,71) 6.052, p=0.016). Furthermore, there is an interaction between Order and Proficiency (F(1,71) 4.053, p=0.048). This interaction points to differential behaviour according to proficiency level on the critical segments for sentences disambiguated by case. Further, there is an interaction of Ambiguity and Language (F(2,71) 8.337, p<0.001). Inspection of the reading times shows that this interaction reflects the slower reading times for sentences disambiguated by verbal agreement by the L1 English and L1 Russian groups.
Table 6.11. Experiment 2: Significant effects for raw and residual reading times.4

<table>
<thead>
<tr>
<th>Segment</th>
<th>Effect</th>
<th>Raw RT</th>
<th>Residual RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3 (NP1)</td>
<td>Order</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Ambiguity</td>
<td>*a</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Ambiguity x Language</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Order x Ambiguity</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>S4 (ADV)</td>
<td>Order x Proficiency</td>
<td>&lt;.1</td>
<td>&lt;.1</td>
</tr>
<tr>
<td></td>
<td>Ambiguity</td>
<td>n.s.  c</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Ambiguity x Language</td>
<td>**</td>
<td>**b</td>
</tr>
<tr>
<td></td>
<td>Order x Ambiguity x Language</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Order x Ambiguity x Proficiency</td>
<td>&lt;.1a</td>
<td>n.s.</td>
</tr>
<tr>
<td>S3&amp;4 (combined)</td>
<td>Order</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Order x Proficiency</td>
<td>a  b  b</td>
<td>a  b</td>
</tr>
<tr>
<td></td>
<td>Ambiguity</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Ambiguity x Language</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Order x Ambiguity</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Order x Ambiguity x Language</td>
<td>&lt;.1a</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>Order x Ambiguity x Proficiency</td>
<td>n.s.</td>
<td>&lt;.1a</td>
</tr>
<tr>
<td>S5 (NP2)</td>
<td>Ambiguity</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Ambiguity x Language</td>
<td>&lt;.1a</td>
<td>&lt;.1a</td>
</tr>
<tr>
<td>S6 (V-part)</td>
<td>Order</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Order x Ambiguity</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>S7 (V-fin)</td>
<td>Order</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Ambiguity</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Order x Ambiguity</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>Total</td>
<td>Order</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Ambiguity x Language</td>
<td>&lt;.1a</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Note: a indicates that the effect did not reach significance in the analysis of only the correct items. b indicates that the effect reached only marginal significance in the analysis of only the correct items. c denotes that the effect reached significance only in the analysis of the correct items. ***: p < .001; **: p < .01; *: p < .05.

Table 6.11 is ordered as follows: For each segment, it first lists main effects and interactions of the main effects with the between-subjects variables. Subsequently, interactions of the within-subjects effects are listed. Finally, interactions of these interactions with between-subjects variables are listed, if any.

---

4 Table 6.11 is ordered as follows: For each segment, it first lists main effects and interactions of the main effects with the between-subjects variables. Subsequently, interactions of the within-subjects effects are listed. Finally, interactions of these interactions with between-subjects variables are listed, if any.
Segment 5, the region of the second NP, shows a main effect of *Ambiguity* (F(1,71) 6.399, p=0.014), which reflects the longer reading times for OS orders disambiguated by case versus sentences disambiguated by verbal agreement. On segment 6, there is a main effect of *Order* (F(1,71) 9.347, p=0.003) as well as an interaction of *Order* and *Ambiguity* (F(1,71) 7.562, p=0.008), which reflects the slower reading times of OS sentences disambiguated by case; this effect seems to spill over from the previous segment, where nominative case on the second NP confirms the OS order of the sentence.

On the sentence-final segment, segment 7, there are main effects of *Order* (F(1,71) 36.537, p<0.001) and *Ambiguity* (F(1,71) 44.455, p<0.001) as well as an interaction of *Order* and *Ambiguity* (F(1,71) 11.638, p<0.001). This interaction shows that reading times increase significantly more for orders disambiguated by verbal agreement on segment 7 than for orders previously disambiguated by case marking.

Given the interactions with the between-subjects factors *Language* and *Proficiency*, the reading times will be further analysed. Note that the interactions of the factor *Language* are with the factor *Ambiguity*, suggesting that (some) non-native groups show differences in reading times depending on the type of disambiguation. It can be seen from scanning the reading times (Appendix F) that the L1 Dutch groups read ambiguously case-marked NPs of either order considerably faster than their L1 English and L1 Russian counterparts. This effect most likely reflects the greater lexical proximity between German and Dutch. In contrast, the interactions of the factor *Proficiency* are with the factor *Order*, suggesting that the proficiency groups show differences in reading slowdowns between SO and OS orders, i.e. they show differences in syntactic (re)analysis. Since we are interested in group differences pertaining to effects of *Order*, I analyse the reading times by proficiency group and not by language group. In the separate analyses of the reading times, I concentrate on the critical segments, for which these interactions were observed, i.e. Segment 3&4 (NP1 & ADV) and segment 7 (V-fin).

**Natives**

On Segment 3&4, the native group demonstrates a main effect of *Order* (F(1,19) 19.751, p<0.001; F(2,17) 5.467, p=0.028), no effect of *Ambiguity* (F(1,19) 2.373, p=0.140) and no interaction between *Order* and *Ambiguity* (F(1,19) 1.029, p=0.323). Pairwise comparisons between (8a) and (8b) show a significant difference in the analysis by subjects and a near-significant effect for items (F(1,19) -3.407, p=0.003; F(2,11) -2.581, p=0.026). This finding shows that OS sentences disambiguated by case are read significantly more slowly than their SO counterparts on the disambiguating segments (Figure 6.6). Figure 6.6 also illustrates that these slowdowns for OS sentences disambiguated by case are replicated towards the end of the sentence.

---

5 Recall from Section 6.6.8 that alpha levels were adjusted for the post-hoc comparisons.
For the sentence-final segment, segment 7, the native group displays main effects of Order (F(1,19) 13.769, p=0.001; F(1,123) 15.127, p=0.001), Ambiguity (F(1,19) 7.775, p=0.012) as well as a marginal interaction between Order and Ambiguity (F(1,19) 4.297, p=0.052). Subsequent pairwise comparisons show a significant difference between (9a) and (9b) (F(1,19) -3.505, p=0.005). This difference attests that OS orders disambiguated by verbal agreement evince longer reading times on the disambiguating segment than their SO counterparts.

**Advanced Group**

For Segment 3&4, the advanced group shows no main effects of Order (F(1,27) 0.288, p=0.596; F(1,69) 0.428, p=0.515) or Ambiguity (F(1,27) 1.040, p=0.317). However, there is an interaction between Order and Ambiguity (F(2,27) 5.979, p=0.021). There is no interaction of Order and Language (F(1,27) 0.502, p=0.611; F(2,69) 0.188, p=0.829). However, the interaction between Order, Ambiguity and Language reaches significance (F(2,27) 4.270, p=0.024) as the consequence of the interaction of Ambiguity and Language that reflects the faster reading times of NPs by the L1 Dutch group. Subsequent pairwise comparisons unearth only a marginal difference between (8a) and (8b) in the analysis by subjects (F(1,29) -2.324, p=0.027; F(1,11) -1.298, p=0.221).
sum, then, the advanced group does not show a robust slowdown on the critical disambiguating segment for OS orders disambiguated by case. Figure 6.7 graphs the reading times for the advanced group.

![Figure 6.7](image-url)

**Figure 6.7.** Experiment 2: Reading times for the advanced group.

On segment 7, the advanced group demonstrates a main effect of Order in the analysis by subjects and marginally by items (F(1,27) 7.959, p=0.009; F2 (1,69) 3.568, p=0.063). There is also a main effect of Ambiguity (F(1,27) 20.094, p<0.001), yet no interaction of Order and Ambiguity (F(2,27) 0.767, p=0.389). Further, no interactions of Order and Language (F(1,27) 0.398, p=0.676; F2 (2,69) 0.106, p=0.900) or Order, Ambiguity and Language (F(2,27) 0.101, p=0.904) arise.

Subsequent pairwise comparisons of (9a) and (9b) on segment 7 show a marginal difference in the analysis by subjects (F(1,29) -2.222, p=0.034; F2 (1,11) -1.291, p=0.223). This difference is not at variance with the difference on segment 7 between (8a) and (8b) (F(1,29) -2.506, p=0.018; F2 (1,11) -1.557, p=0.148). Thus, the advanced group does not evince a robust locally specific slowdown on the disambiguating segment for OS orders disambiguated by verbal agreement; rather, the advanced group shows a general end-of-sentence slowdown for OS orders.
Near-Native Group

On Segment 3&4, the near-natives evince a main effect of Order ($F_{1}(1,25) 14.726$, $p=0.001$; $F_{2} (1,69) 7.815$, $p=0.007$) and an interaction of Order and Ambiguity ($F(1,25) 25.642$, $p<0.001$). There is a marginally significant main effect of Ambiguity ($F(1,25) 2.990$, $p=0.096$) as well as an interaction of Ambiguity and Language ($F(2,25) 4.399$, $p=0.023$).

![Graph showing reading times for the near-natives](image)

Figure 6.8. Experiment 2: Reading times for the near-natives.

Again, this interaction arises from the comparatively faster reading times of the L1 Dutch group for ambiguously case-marked NPs. Further, there are no interactions of Order and Language ($F_{1}(2,25) 1.313$, $p=0.287$; $F_{2} (2,69) 0.920$, $p=0.403$) or Order, Ambiguity and Language ($F(2,25) 0.641$, $p=0.535$). Subsequent pairwise comparisons between (8a) and (8b) reach significance ($F_{1}(1,27) -6.271$, $p<0.001$; $F_{2} (1,11) -4.178$, $p=0.002$). In line with the native speakers, then, the near-natives display robust incremental slowdowns for the OS order on the region of disambiguation (see Figure 6.8).

For segment 7, there is a main effect of Order ($F_{1}(1,25) 14.827$, $p=0.001$; $F_{2}(1,69) 21.125$, $p<0.001$) and Ambiguity ($F(1,25) 21.015$, $p<0.001$), as well as an interaction between Order and Ambiguity ($F(1,25) 7.353$, $p=0.012$). The interactions of Order and Language ($F_{1}(2,25) 0.237$, $p=0.791$; $F_{2} (2,69) 0.120$, $p=0.887$) and Order,
Ambiguity and Language (F(2,25) 0.640, p=0.536) fail to reach significance. In pairwise comparisons, the difference between (9a) and (9b) becomes highly significant (F(1,27) -3.643, p=0.001; F₂ (1,23) -3.879, p=0.003). By contrast, the comparison between (8a) and (8b) is not significant (F(1,27) -2.151, p=0.041; F₂ (1,11) -1.644, p=0.128). This finding demonstrates that the slowdowns on OS orders on segment 7 are specific to disambiguation by verbal agreement for the near-natives. In sum, like the native speakers, the near-native L2 group evinces locally specific reading slowdowns for OS orders on the regions of disambiguation.

To sum up, analysing the reading data according to proficiency level, we find that the natives and the near-natives show isomorphic behaviour, whereas the advanced group fails to evince robust slowdowns on the critical segments for either type of syntactic disambiguation. These findings hold across L1s.

6.6.11. Discussion

In Experiment 2, the following results were found:

- In comprehension accuracy, all groups show a strong preference for SO orders over OS orders.
- The comprehension accuracy for OS orders disambiguated by case is significantly higher than for OS orders disambiguated by verbal agreement. This points to the differential status of case and verbal agreement.
- All groups show similar patterns in comprehension accuracy with a few exceptions:
  o Although they show the same relative pattern of accuracy on OS orders, the advanced L2 groups demonstrate lower accuracy than the near-native L2 groups.
  o The L1 Dutch groups do not show an asymmetry in accuracy between OS orders disambiguated by case and OS orders disambiguated by verbal agreement.
- Response times for the comprehension questions are longer for OS than for SO orders for all groups.
- The reading times show a slowdown on Segment 3&4 for OS orders disambiguated by case, as indicated by the interaction of Order and Ambiguity. There is an interaction between Order and Proficiency, yet not between Order and Language, indicating that the groups behave differently according to proficiency level, yet not according to language.
- The reading times show a slowdown on segment 7 for OS orders disambiguated by verbal agreement, as indicated by the interaction of Order and Ambiguity.
- Analyses of the reading times by proficiency groups show that the native and the near-natives behave similarly in evincing locally specific reading slowdowns for OS orders on Segment 3&4 (for case disambiguation) and on segment 7 (for verbal agreement disambiguation).
- In the analysis of the reading times, the advanced learners do not show robust slowdowns for OS orders disambiguated by case on Segment 3&4. Furthermore, the advanced group does not show a specific slowdown on segment 7 for OS orders disambiguated by verbal agreement.

**L1 effects in comprehension accuracy**

In comprehension accuracy, the results of Experiment 2 bear out the strong subject-first preference for all groups. This finding is backed up by delays for OS orders in the response times to the comprehension questions. Comprehension accuracy is further affected by proficiency level; yet, there are no L1 effects on either type of scrambled OS order. Irrespective of whether or not the L1 instantiates scrambling or case marking, the near-natives converge on native levels of comprehension accuracy.

Further, the L1 English and L1 Russian groups demonstrate an asymmetry in comprehension accuracy according to type of disambiguation. The L1 Dutch group fails to perform differentially by type of disambiguation: The L1 Dutch groups attain levels of comprehension accuracy for OS orders disambiguated by verbal agreement on a par with the other non-native groups; yet, they score much lower on OS orders disambiguated by case. In the reading time data, however, the L1 Dutch groups pattern alongside their proficiency-matched L1 English and L1 Russian counterparts.

It is unclear why the L1 Dutch groups should show this specific behaviour different from the L1 English groups. Note that the L1 Dutch groups read the sentences at a faster pace than the other non-native groups at the same proficiency levels. Part of the low comprehension accuracy by the L1 Dutch group might accordingly reflect a speed-accuracy trade-off. On this account, faster reading might entail lower sensitivity to morphosyntactic features, such as case and verbal agreement, and lead to depressed comprehension accuracy on non-canonical orders. However, on such a view, (a) one would expect the Dutch groups to perform worse on all OS orders and not exclusively OS orders disambiguated by case marking and (b) one would expect the low comprehension accuracy to be mirrored in failure to use case marking and verbal agreement in on-line reading. However, the L1 Dutch group does not behave any differently from the proficiency-matched L1 English or L1 Russian speakers in reading patterns, despite the fact that the L1 Dutch group shows faster reading times. It is therefore unlikely that the faster reading speed of the L1 Dutch groups causes their lower comprehension accuracy. We can further consider this aspect in the following experiment: Experiment 3, a speeded-grammaticality judgement task, will show whether the group differences in
comprehension accuracy persist in a task that imposes uniform processing pressures on all groups.

**L1 effects in reading times**

The statistical analysis does not elicit interactions with the factor *Language* that index relevant differences between the L1 groups in terms of their reading patterns. At the advanced level, the fact that Russian instantiates pre-subject scrambling flagged by morphological case cues does not confer L1 Russian speakers a processing advantage for scrambling over L1 English and L1 Dutch speakers. Conversely, at near-native levels, the fact that English and Dutch do not instantiate pre-subject scrambling flagged by morphological case cues does not confer a processing disadvantage on the L1 English and the L1 Dutch groups compared to the L1 Russian speakers. In this way, the findings on processing filler-gap dependencies in scrambling concur with previous research on the L2 processing of *wh*-movement that did not report L1 differences (Marinis et al., 2005; Williams, 2006; Williams et al., 2001).

**Proficiency effects in reading times**

Rather, reading patterns differed according to proficiency. The advanced groups do not show robust effects of incremental phrase-structure revision according to morphological cues. The failure to show on-line reanalysis effects is not contingent on the type of disambiguation; even for disambiguation by verbal agreement, which is realized across all the languages investigated, advanced learners do not evince specific slowdowns.

Since this effect generalizes across L1s and morphosyntactic information types, it furnishes tentative evidence that lower-proficiency learners do not execute incremental reanalysis according to morphosyntactic cues while reading sentences in the L2. Rather, the advanced groups appear to rely on linear order in syntactic function assignment and to assign the first NP the role of the grammatical subject and the second NP the role of the grammatical object, regardless of case marking or verbal agreement.

Note, however, that there are no robust interactions with the factor *Proficiency* in total sentence reading times or in the response times to the comprehension questions. Hence, the advanced groups do in fact show slowdowns for OS orders at the end of sentences — like the near-native and native groups. This combination of findings suggests that, rather than incrementally in the course of reading the sentences, the advanced groups perform reanalysis at a later stage in sentence processing or post-processing after having read sentences in their entirety.

The near-natives show incremental reanalysis effects like the native speakers. As shown by the interactions of the factors *Order* and *Ambiguity* in the reading delays, reanalysis in the near-natives is locally specific to the regions of syntactic
disambiguation, i.e. Segment 3&4 for case and segment 7 for verbal agreement. Target-like processing of morphosyntax thus seems to emerge as a function of proficiency in the adult L2 processing of syntactic dependencies.

Hence, not only do the near-natives come to acquire target-like grammatical knowledge of scrambling (Experiment 1), they can also use this knowledge rapidly in processes of on-line phrase-structure building and syntactic reanalysis. This obtains for a syntactic phenomenon not all L1s realize (scrambling) and also for morphological cues not all L1s realize (case). Experiment 2 thus yields evidence that the integration of morphological and syntactic knowledge can be executed automatically and efficiently in highest-proficient L2 speakers under experimental conditions approximating real-time reading comprehension. The present findings resonate with research across a range of processing phenomena that attests that differences in proficiency lead to substantial processing changes (e.g. Abutalebi et al., 2001) and that L2 speakers at the top end of L2 proficiency converge on native speaker processing performance (Rossi et al., 2006).

In sum, Experiment 2 shows similarities between near-native and native processing as well as differences between advanced L2 and native processing. To complete the investigation of the extent to which L2 processing of the morphosyntax of scrambling conforms to native patterns, Experiment 3 probes the processing of case and verbal agreement in ungrammatical sentences.

### 6.7. Experiment 3: Speeded grammaticality judgements

Speeded grammaticality judgements have been widely used to tap processing strategies. Even though speeded judgement tasks are, strictly speaking, off-line tasks since they measure the outcome of sentence processing, the speeded presentation of the stimuli and the rapidly enforced judgement are taken to reflect processing strategies because the pace of the task (a) forces the parser to adopt its preferred parsing route and (b) does not allow for enough time to complete reanalysis (e.g. Bader & Bayer, 2006). The rationale underlying the speeded judgement paradigm is that, under time pressure, sentences dispreferred by the parser elicit lower accuracy scores and higher reaction times than comparable control sentences. The speeded grammaticality judgement task aims to elicit the following target pattern (P.3).

(P.3) **Target pattern for Experiment 3**

(a) Scrambled OS orders show decreased accuracy and longer response latencies compared to SO orders.

(b) Sentences ungrammatical by verbal agreement are judged more accurately than sentences ungrammatical by case marking.
The contrast between case marking and verbal agreement reflects different degrees of grammatical informativity of these features for syntactic reanalysis (e.g. Bader & Bayer, 2006; Fodor & Inoue, 2000). Briefly, ungrammatical case marking erroneously signals potential reanalysis to the parser (i.e. from SO to OS). It thus gives rise to false positive responses to ungrammaticalities, at least for judgements under time pressure. In contrast, ungrammatical verbal agreement does not provide information as to how to rectify a parse, so that it is judged as ungrammatical more confidently (see Table 6.7).

In light of this target processing pattern, the research questions for the L2 groups are as follows.

(Q3.1) Do L2ers show effects of syntactic reanalysis for scrambling in sentence-level reaction times under speeded presentation?
(Q3.2) Do L2ers make a difference in judgement accuracy under speeded presentation between case violations and verbal agreement violations?
(Q3.3) Are there L1 differences or proficiency differences in speeded judgements?

Further, Experiment 3 allows for direct group comparisons in a task that holds the processing demands, i.e. the speed of presentation, constant across groups, such that potential differences in the speed or automaticity of the coordination of morphological and syntactic knowledge can be observed.

6.7.1. Materials

In a factorial design, 36 sixtuplets were created: Sentences differed according to the factor **Type** (grammatical, case violation, agreement violation). Sentences within each sixtuplet were in either SO or OS order (Factor **Order**). For case violations, the double nominative in (14) was the ungrammatical counterpart of the SO order (12) and double accusative in (15) was the ungrammatical equivalent of the OS order (13). Two additional versions within each sixtuplet set were constructed by reversing the position of the nouns (i.e. N1-N2 and N2-N1), so that any potential effect of lexical semantics or pragmatics of the SO and OS manipulation would be completely matched. This reversal was not included as a factor in the statistical design. The speeded grammaticality judgement task thus included the six following conditions: Grammatical SO (12) and OS (13) sentences disambiguated by case marking versus case violations, i.e. ‘doubly-nominative marked’ sentences *(der-der)* (14) and ‘doubly-accusative marked’ sentences *(den-den)* (15), and violations of verbal agreement (number) in SO order (16) and in OS order (17). All experimental items for Experiment 3 are given in Appendix G.
(12) Er glaubt, dass der Bäcker seit langer Zeit den Metzger beliefert hat. 
He believes that the baker for long time the butcher supplied has

(13) Er glaubt, dass den Bäcker seit langer Zeit der Metzger beliefert hat. 

(14) *Er glaubt, dass der Bäcker seit langer Zeit der Metzger beliefert hat. 
He believes that the baker for long time the butcher supplied has

(15) *Er glaubt, dass den Bäcker seit langer Zeit den Metzger beliefert haben. 
He believes that the butcher supplied have

(16) *Er glaubt, dass der Bäcker seit langer Zeit den Metzger beliefert haben. 

(17) *Er glaubt, dass den Bäcker seit langer Zeit der Metzger beliefert haben. 

6.7.2. Plausibility and reversibility

The plausibility and reversibility of items was tested in a separate off-line study that was equivalent in design to the plausibility study for Experiment 2 and was administered to the same 16 participants.

6.7.3. Assignment to groups and lists

Group and list assignment was as for Experiment 2.

6.7.4. Fillers

In addition to the experimental items, the task included 66 fillers, thus yielding a total of 102 items. The fillers were the same for each list and encompassed various constructions. In order to prevent the subjects from developing response strategies based on surface generalizations, the fillers included grammatical dative SO sentences as in (18) which contained two NPs with the determiner der to counterbalance the ungrammatical double-nominative violations (14).

(18) Er sagt, dass der Gärtner der Bäuerin geholfen hat. 
He says that the gardener the farmer helped has

‘He says that the gardener helped the farmer.’
To counterbalance the ungrammatical accusative den-den construction (15), double object constructions were used as in (19) that exhibited a grammatical den-den sequence.

(19) Gestern hat den Unternehmer den Ehrengästen die Präsidentin vorgestellt.
Yesterday has the ACC entrepreneur the DAT honorary guests the president introduced
‘Yesterday, the president introduced the entrepreneur to the the honorary guests.’

To counterbalance the verbal agreement violations in (16), which combine two singular NPs with a plural auxiliary, the grammatical plural sentences in (20) were devised that combined two plural NPs with a plural auxiliary. The remaining fillers were of various other structures.

(20) Ich glaube, dass die Gerüstbauer am Freitag die Bauarbeiter düpiert haben.
I believe that the scaffolders on Friday the building workers embarrassed have
‘I believe that the scaffolders embarrassed the building workers on Friday.’

6.7.5. Order of presentation

The order of experimental items and fillers was randomized. The order of sentences was randomized automatically for each participant at the beginning of the experimental session.

6.7.6. Procedure

Participants were tested individually. They sat in front of a laptop computer with a 15-inch TFT screen. The keyboard was covered with a blind that left only the spacebar and two keys visible. The spacebar was the ‘Go’ key. The two other keys were slightly offset from the rest of the keyboard on the bottom right. The left key was marked with a green sticker (the ‘yes’ key) and the right key was marked with a red sticker (the ‘no’ key).

A trial was initiated by pressing the ‘Go’ key. Each trial sentence was preceded by a fixation star in the centre of the screen. When the participant pressed the ‘Go’ key, the sentences were presented word-by-word in the centre of a 15-inch TFT screen in white font (Courier New, 18) against a black background. Following similar speeded judgement tasks on native speakers (Meng & Bader, 2000; Schlesewsky, Fanselow et al., 2003), the rate of presentation was 250 ms per word plus 17 ms per letter. There was no inter-stimulus interval. Sentences were presented without punctuation. After the final word of each sentence, the screen changed colour and the participants made an immediate binary grammaticality judgement by pressing the green (‘grammatical’) or red (‘ungrammatical’) button. Once the response had been given, the screen changed to display the fixation star, and the following trial could be initiated by pressing the space
Participants’ responses and response times following the off-set of the final word in the sentence were recorded.

The participants were told that the experiment was about reading comprehension in a (second) language and were instructed to judge the sentences as accurately and as quickly as possible. They were told that the sentences contained no punctuation and that there were no spelling mistakes. The presentation of the actual items was preceded by a page of instructions and six practice items. Feedback on the participants’ responses in the practice items was given by the experimenter. Once the participants had understood the procedure and could operate the buttons, the actual experiment was started. The participants did not receive any feedback during the experimental session. Most participants completed the task within 12 minutes.

6.7.7. Analysis and results

Response times of a participant for a sentence above or below two standard deviations of the group mean for that sentence were trimmed to the group mean of the sentence plus or minus two standard deviations, respectively. This affected less than 2% of the trials in each group.

Three Repeated Measures ANOVAs with the factors Order (SO, OS) or Type (grammatical, case violation, agreement violation) as within-subjects factors and Language (German, English, Dutch, Russian) and Proficiency (native, advanced L2 and near-native L2) as between-subjects factors were performed on accuracy scores and on the reaction times. In subsequent pairwise comparisons, the significance level was set to an alpha level of .025. Analyses of the reaction times were run separately on all items and on only those items that had been judged correctly.

Table 6.12 displays the number (percent) of accurate responses per condition. Mean response times (RT) for all responses are also given.
### Table 6.12. Experiment 3: Accuracy (in percent) and Response Times in Speeded Grammaticality Judgement Task.

6.7.7.1. Accuracy

The between-groups comparison for sentences (12) and (13) shows a significant main effect of Order (F(1,72) 23.073, p<0.001). There are no significant interactions with the factors Language or Proficiency, bearing out that all groups demonstrate an SO preference compared to the OS order.

Let us turn to the use of case marking in speeded judgements. To determine whether participants make a difference between grammatical and ungrammatical case marking, comparisons of accurate judgements on (12) and (13) (Type: grammatical), on the one hand, with the false positive judgements of (14) and (15) (Type: case violation),
on the other hand, were run. They reveal a main effect of Order \((F(1,72) = 11.105, p=0.001)\) and Type \((F(1,72) = 132.255, p<0.001)\) as well as an interaction between Order and Type \((F(1,71) = 8.444, p=0.005)\). There are also interactions between the factors Type and Language \((F(2,72) = 3.926, p=0.024)\) and Type and Proficiency \((F(2,72) = 21.887, p<0.001)\). This interaction suggests that not all proficiency groups distinguish between grammatical and ungrammatical case marking.

Subsequent pairwise analyses by proficiency group reveal highly significant differences for the natives and the near-natives between grammatical SO and ungrammatical double nominative orders (natives: \(F(1,19) = 9.158, p<0.001\); near-natives: \(F(1,28) = 6.620, p<0.001\)) as well as between grammatical SO and ungrammatical double accusative orders (natives: \(F(1,19) = 5.013, p<0.001\); near-natives: \(F(1,28) = 5.298, p<0.001\)). In contrast, while the advanced group also shows a significant difference between grammatical SO and ungrammatical double nominative orders \((F(1,29) = 2.504, p=0.018)\), they display only a marginally significant difference between grammatical OS and ungrammatical double accusative orders \((F(1,29) = 2.340, p=0.026)\). This finding shows that the advanced group distinguishes between grammatical and ungrammatical case marking less robustly than the near-natives and the natives do.

Finally, consider whether case ((14) & (15)) and number ((16) & (17)) violations are treated differently. The comparison of case and number violations yields a main effect of Type \((F(1,72) = 137.185, p<0.001)\) and an interaction of Type with the factor Proficiency \((F(2,72) = 11.857, p=0.001)\). This interaction reflects the much lower accuracy in detecting the ungrammaticality of case violations by the advanced group. Subsequent analyses by proficiency group yield highly significant effects of Type for each non-native proficiency group (near-natives: \(F_1(1,28) = 24.786, p<0.001\); \(F_2(1,11) = 57.321, p<0.001\); advanced: \(F_1(1,29) = 125.827, p<0.001\); \(F_2(1,11) = 112.729, p<0.001\)), bearing out that, for each group, the judgement accuracy on detecting case violations is lower than on detecting verbal agreement violations. Note that the high accuracy on ruling out verbal agreement violations for all groups also shows that all groups can in principle make distinctions between grammatical and ungrammatical sentences under the speeded conditions of the task. It is also worth noting that the L1 Dutch groups do not demonstrate different behaviour in judgement accuracy of case versus verbal agreement. The between-language effects in Experiment 2 may thus at least partially result from differences in reading speed in the self-paced task, since the effect is not replicated in the speeded judgement task that holds processing speed constant.

6.7.7.2. Reaction times

The between-groups analysis of the SO (12) and the OS (13) orders yields a significant main effect of Order in the analysis of all items \((F(1,77) = 21.069, p<0.001)\) and in the
analysis of correct trials (F(1,70) 18.534, p<0.001), yet no interactions with the factors Language or Proficiency.

Nevertheless, in light of the differences between proficiency groups in the accuracy scores, I ran subsequent ANOVAs by proficiency group. These ANOVAs replicate the main effect of Order for the natives (all items: F₁(1,19) 8.777, p=0.008; F₂(1,11) 6.966, p=0.023; correct trials: F₁(1,18) 6.893, p=0.017; F₂(1,11) 3.677, p=0.081) and the near-natives (all items: F₁(1,28) 11.110, p=0.002; F₂(1,11) 16.210, p=0.002; correct trials: F₁(1,28) 6.142, p=0.019; F₂(1,11) 4.015, p=0.070). For the advanced group, however, there is no significant effect of Order in the analysis of all items (F₁(1,29) 3.854, p=0.059; F₂(1,11) 2.021, p=0.183); in the analysis of the correct trials, the main effect of Order becomes significant in the analysis by subjects (F₁(1,28) 5.690, p=0.024; F₂(1,11) 2.629, p=0.133). The separate group analyses document that the advanced group does not robustly evince slowdowns for OS orders vis à vis SO orders. All other comparisons do not yield significant effects (p>0.05) and are hence not reported.

6.7.8. Discussion

Experiment 3 yielded the following results:

- All groups show an SO preference as documented in the decreased judgement accuracy on OS orders.
- For all groups, judgement accuracy on verbal agreement violations is higher than judgement accuracy on case violations.
- The non-native groups’ behaviour differs according to proficiency level as well as L1.
  o The advanced groups do not show robustly longer reaction times to OS orders that would be indicative of reanalysis.
  o The advanced groups score below chance on detecting case violations, whereas the near-natives perform significantly better.
  o The advanced groups do not make robust distinctions between grammatical and ungrammatical case marking.
  o At near-native levels, the L1 Russian group attains higher accuracy in detecting ungrammatical case marking than the L1 English and L1 Dutch groups.

Experiment 3 confirms that case and verbal agreement behave differently in processing. Finding that verbal agreement violations are judged more accurately than case violations confirms that the difference in garden-path strength between the disambiguation of

---

6 Note that for both the natives and the near-natives, the main effect of Order only reaches marginal significance in the analysis by item when only correct trials are considered.
sentences by case versus by verbal agreement in Experiment 2 is not due to a difference in the linear position of the regions of disambiguation (see also Meng & Bader, 2000). Otherwise case violations, for which information identifying the violation occurs earlier than for verbal agreement violations in Experiment 3, should receive higher accuracy scores than verbal agreement violations. Hence, the asymmetric decrease of judgement accuracy on case versus verbal agreement under speeded presentation points to differential properties of these morphosyntactic features for reanalysis.

Significant differences between case disambiguation and disambiguation by verbal agreement obtain for all groups, irrespective of L1 or proficiency differences. Accordingly, all groups treat case and verbal agreement differently in the processing of ungrammatical sentences.

**Proficiency**

Effects of proficiency surface, first, in the response times between SO and OS orders and, second, in accuracy of judgements on case violations. First, unlike the near-natives, the advanced speakers do not evince robust response-time slowdowns for OS orders. Second, unlike the near-natives, all advanced groups uniformly accept case-marking violations at levels well above chance. No group differences obtain depending on whether the L1 employs case marking (Russian) or not (English, Dutch). Experiment 2 suggested that advanced speakers across L1s do not effect incremental reanalysis according to morphological information. The findings of Experiment 3 support this interpretation in that (a) the advanced groups do not display reaction-time evidence of syntactic reanalysis, and (b) their use of case marking is subject to almost complete breakdown under time pressure.

**L1 effects**

Since all advanced group perform similarly, Experiment 3 bolsters the findings from Experiment 2 that the processing of OS orders or case marking is not subject to L1 influence at the advanced level. At near-native levels, however, L1 effects surface in that the L1 Russian group outperforms the L1 English and L1 Dutch groups. In fact, the L1 Russian group patterns with the natives, while the L1 English and L1 Dutch groups only attain levels of ungrammatical case detection indistinguishable from chance.

Unlike the previous experiments that allow subjects to process sentences at their own pace, Experiment 3 imposes the same computational demands on all subjects. Under these conditions, the convergence attested across the near-native groups in Experiment 2 gives way to differential performance according to L1. While no between-group differences obtain for violations of verbal agreement (realized across L1s), case-marking violations do give rise L1-specific performance. Experiment 3 thus finds a ‘delayed L1
effect’ in that facilitatory effects of L1-TL similarity in case marking surface at the highest levels of proficiency in speeded L2 processing, whereas L2 speakers at lower levels of proficiency do not seem to benefit from analogous L1 properties.

In sum, Experiment 3 shows that the processing of ungrammatical sentences in the L2 resembles native patterns with respect to asymmetries between case and verbal agreement, reflecting their different grammatical status in reanalysis. The speeded task also replicates the proficiency effect on target reanalysis reported in Experiment 2. It further uncovered specific L1 differences for processing at increased speed, suggesting that convergence in processing morphosyntax in the L2 is affected by L1 processing routines.

6.8. General discussion of Experiments 1-3

This chapter explored the syntax-morphology interface at L2 ultimate attainment. The target morphosyntax in the context of scrambling is expressed in a complex processing pattern. As a starting point for the discussion, this pattern of native processing of the morphosyntax is repeated in (21).

(21) Native processing of the morphosyntax of subject-object ambiguities

(a) Native processing is guided by a structural subject-first preference that reflects the operation of economy-based parsing principles in incremental phrase-structure building (e.g. the Minimal Chain Principle, DeVincenzi, 1991).

(b) OS worders invoke syntactic reanalysis. Reanalysis is measurable in global as well as local reading delays at the points of morphological disambiguation to the OS order.

(c) In grammatical sentences, OS orders disambiguated by number marking on the verb evince greater processing cost than OS orders disambiguated by case marking, while the inverse holds in ungrammatical sentences (see Table 6.7). This contrast reflects different grammatical informativity of morphosyntactic features for syntactic reanalysis (Fodor & Inoue, 2000; Meng & Bader, 2000).

Three experiments probed (a) the extent to which L2 processing of morphosyntax conforms to this pattern of native parsing and (b) the extent to which L1 differences and proficiency level affect L2 processing. The results show (a) differences in performance between off-line and on-line tasks, (b) between-group differences according to proficiency level and (c) between-group differences according to L1. Table 6.13 gives a schematic overview of the results for each group in terms of convergence on the native target pattern. Since most group differences in Experiments 1 to 3 aligned with proficiency differences, Table 6.13 is organised along the proficiency dimension.
Table 6.13. Overview of the results from Experiment 1-3. ‘+’ denotes convergence and ‘-’ denotes non-convergence.

Only the L1 Russian near-natives converge on native performance on morphosyntax in all aspects, including the detection of case marking violations at elevated speed. At the same time, all L2 near-native groups demonstrate convergence on the target pattern of morphosyntax in L2 processing. Across Experiments 1 to 3, all near-natives evince target interactions of morphology and syntax for reanalysis in L2 processing.

The reanalysis pattern of the near-natives in Experiments 2 and 3 is fully compatible with native-language parsing models which posit that native speakers (a) employ a phrase-structural preference like the Minimal Chain Principle (DeVincenzi, 1991), (b) establish syntactic relations incrementally in parsing, and (c) attempt reanalysis by means of syntactic relations in parsing. The different degrees of grammatical informativity of case features and verbal agreement features for reanalysis lead to a disjunction in parsing accuracy between dispreferred OS sentences and ungrammatical sentences (Table 6.7). Since near-natives show evidence of exactly this performance, syntactic parsing principles and morphosyntactic features are accessed and used in a target-like way in processing the L2.
These findings are in conflict with previous research findings on incremental processing of L2 morphosyntax. In reading-time studies, e.g. Jiang (2004; 2007) finds no incremental slowdowns in the processing of verbal agreement violations for advanced, though not near-native, L1 Chinese learners of English. At the same time, the current findings echo ERPs studies on L2 morphosyntax, where high-proficient L2 speakers showed target-like neurophysiological responses to morphosyntactic violations (Friederici et al., 2002; Hahne et al., 2006; Osterhout et al., 2006; Rossi et al., 2006). Experiments 1-3 thus attest that evidence of the incremental and automatic use of inflectional morphology emerges only at the highest levels of proficiency.

Indeed, the relatively lower-proficient advanced L2 groups do not evince fully target-like patterns of morphosyntax. In the processing experiments, however, the differences in L2 processing performance between advanced L2ers and near-natives seem to be a matter of degree. With reference to the native processing pattern summarized in (21), the L2 groups, including the advanced learners, demonstrate analogous processing patterns in that they all show (a) evidence of a subject-first preference, (b) global reading time differences between SO and OS orders signaling reanalysis, and (c) differences in parsing accuracy depending on syntactic feature and sentence type (Table 6.7). Given that all lexical, semantic, pragmatic and frequency-based cues to sentence interpretation had been removed in Experiments 2 and 3, the comprehension patterns indicate that the L2 groups are sensitive to morphosyntactic information in parsing; otherwise, no difference in the treatment of SO or OS orders would be expected, since the OS orders can felicitously be interpreted as SO orders if morphological case or number marking is simply disregarded. Yet, even the advanced groups (a) comprehend a significant proportion of OS orders accurately, although the only disambiguating cue is case or verbal agreement information, and (b) manifest an interaction between grammaticality and morphosyntactic feature type as in Table 6.7. These results imply, respectively, (a) that morphosyntactic features can be used, and (b) that the grammatical status of morphosyntactic features does inform phrase-structural reanalysis. Nevertheless, the advanced L2ers have manifest problems using morphosyntactic information incrementally in L2 processing in self-paced reading and speeded judgements.

L1 effects

For incremental effects of morphosyntax (Experiment 2), there are no L1 differences between groups. These findings are in line with research on wh-movement in high-intermediate to advanced L2 speakers of different L1 backgrounds by Marinis et al. (2005), Williams et al. (2001) and Williams (2006) that do not report processing differences according to L1s. Marinis et al. (2005) report that advanced L2 speakers do not evince native-like incremental effects of phrase-structure building in long-distance wh-movement, irrespective of whether the L1 has overt wh-movement. The non-
convergence attested for the advanced groups in Experiments 2 and 3 suggests that problems with the on-line use of morphosyntax in L2 processing extend to syntactic reanalysis in subject-object ambiguities, again irrespective of whether the L1 has case and/or verbal agreement morphology. In contrast to previous studies that found that L2ers across L1s did not use morphosyntax in a target manner, though, Experiment 2 shows that L2ers across L1s do effect target syntactic reanalysis at near-native levels.

In Experiments 1 and 3, L1 effects are related to L1 instantiation of case marking, since the L1 Russian groups differ from the other L1 groups. At advanced levels, the L1 Russian group demonstrates target-like knowledge of case marking off-line, unlike the L1 English and L1 Dutch advanced groups; however, this difference does not translate to processing (Experiments 2 & 3). In Experiment 3, the L1 Russian near-natives outperform their L1 English and L1 Dutch counterparts in detecting case violations in speeded judgements. In both cases, these L1 differences seem to relate to the proceduralization of grammatical knowledge, rather than to differences in grammatical knowledge per se. The L1 Russian advanced group does not seem capable of applying morphosyntactical knowledge incrementally in processing, whereas the L1 Russian near-natives are demonstrably better than the other L1 groups at applying morphosyntax rapidly in processing.

In sum, Experiments 1-3 find that L2ers converge on native morphosyntax in L2 processing at near-native levels. Target patterns of processing morphosyntax are observed across L1 groups, which indicates that L2ers fully acquire the TL morphosyntax and are capable of using it rapidly and incrementally in sentence processing. These results are incompatible with accounts of L2 acquisition or L2 processing that posit representational impairment in L2 morphosyntax (e.g. Clahsen & Felser, 2006b; Hawkins & Chan, 1997; Tsimpli, 2003). At lower levels, however, L2ers suffer problems with L2 morphosyntax. In addition, L1 effects surface in the degree of use of case morphology in L2 processing in some tasks.

6.8.1. Linking and checking in the L2 processing of scrambling

Let us explore how these differences in group performance according to proficiency, L1, and task can be accounted for within the framework of Bader & Bayer’s (2006) model of Linking and Checking (Chapter 4). The parsing of the morphosyntax of subject-object ambiguities involves, inter alia, (a) the phrase-structural preference for subject-initial orders and (b) morphological checking of case and verbal agreement information to confirm or disconfirm the initial subject-first analysis of the parse. As the time course of native processing shows, these processes apply consecutively. First, the parser links incoming NPs to argument positions established by universal phrase-structure building principles like the Minimal Chain Principle (e.g. DeVincenzi, 1991). Accordingly, the first NP will be linked to the subject position. Second, as per the LBCA, morphological
checking applies to verify this assignment. If, e.g., accusative case marking signals that the first NP cannot be the subject, the phrase structure will be revised (reanalysis).

Morphological checking implicates search of features and checking. Due to syncretism of case, gender and number marking in the German case paradigm (Chapter 3), there is no one-to-one mapping between forms and features. Feature checking thus requires searching the lexicon which, in turn, involves processing effort; hence, its success depends on the time available for its execution in processing tasks and on the automaticity of feature checking. The time available for feature checking is determined by task demands. Further, I assume that the automaticity of feature checking in L2 processing is codetermined by L2 proficiency and practice in executing feature checking. I consider how both dimensions interact in the next section.

6.8.1.1 Task demands and automaticity

Experiments 1 through 3 systematically differ in their task demands in the sense that each imposes different time limits on morphological checking. For Experiments 1 through 3, task demands are lowest in off-line judgements that allow potentially unlimited time for (meta-) linguistic analysis, higher in the self-paced reading task that resembles the speed of real-time reading, and highest in the speeded grammaticality judgement task that forces subjects to make premature judgements. These tasks thus impose different demands on morphological checking by gradually narrowing down the time available for its execution. To start with the off-line judgement task, clear effects of grammatical L1 transfer are attested at advanced levels for word order and case marking. At the near-native level, the higher proficiency of the L2 groups cancels these L1 effects.

However, differences in grammatical knowledge do not translate one-to-one to processing performance. The advanced groups apply the subject-first preference by linking the first NP to subject position; yet they do not seem to execute morphological checking incrementally to verify this linking. Rather, the assignment of the role of grammatical subject to the first NP prevails even in the face of incompatible (accusative) case information on the first NP or incompatible verb information (faulty number agreement) on the sentence-final segment in first-pass parsing. Revision to the OS order, if at all, occurs non-locally across and after reading the sentence.

Since this pattern of non-incremental processing holds for all L2 advanced groups, irrespective of whether they show target knowledge (L1 Russian) or not (L1 English, L1 Dutch) off-line, the uniform on-line behaviour can be related to the increased task demands of the on-line reading task that exceed the capacities of the advanced groups.

In Bader & Bayer’s (2006) model of Linking and Checking, the advanced groups’ performance can be construed as incomplete processing in that linking applies without completed morphological checking (see also Bornkessel & Schlesewsky, 2006a). As a
result, the advanced groups construct canonical subject-first structures, as borne out in their low comprehension accuracy of OS orders. Other psycholinguistic models, discussed in Chapter 2, characterize incomplete processing as the initial processing step in a dual-route architecture. Both in the LAST model of Townsend & Bever (2001) and in the model of ‘good-enough’ processing by Ferreira (Ferreira, 2003; Ferreira & Patson, 2007), sentence processing comprises an initial ‘shallow’ analysis that is always computed before or in tandem with a full parse that includes, e.g., morphological checking. In particular, both models assume that an NVN (noun-verb-noun) strategy is applied to the input, irrespective of morphological marking. The NVN strategy describes a mapping of noun-verb-noun sequences to a thematic proto-agent, action, patient template as a processing heuristic in order to quickly obtain a semantically interpretable output or a ‘good enough’ representation of the sentence. Irrespective of subsequent full parsing, the NVN template of ‘good enough’ processing may partially survive to interpretation, also in native processing (Christianson & Slattery, 2005; Ferreira, 2003).

In Experiment 2, the natives show incremental reanalysis effects indicative of morphological checking, i.e. full parsing. At the same time, the answers to the comprehension sentences show that they correctly interpret OS orders only 56% of the time for disambiguation by case marking and only 28% of the time for disambiguation by verbal agreement. In all other cases, natives incorrectly assign the parse an SO interpretation (see also Meng & Bader, 2000). The use of the NVN template as a processing strategy is thus not restricted to the L2 subjects. Crucially, though, for the natives and the near-natives, ‘good enough’ processing is accompanied by full parsing and morphological checking, as expressed in incremental reanalysis effects, whereas the task demands of self-paced reading restrict the L2 advanced groups to ‘good enough’ processing.

In the speeded judgement task that imposes yet higher computational demands, the use of case marking in ungrammaticality detection — which requires morphological checking — is subject to complete breakdown in the advanced L2 groups and the L1 English and L1 Dutch near-natives. For the natives and the L1 Russian near-natives, the selectively decreased accuracy on case violations also shows that the task demands exceed computational capacities for morphological checking.

Consider the L1 effects in Experiment 3 for the near-natives. Given that all near-natives performed target-like in self-paced reading (Experiment 2), the L1 effects in Experiment 3 index L1 differences in the speed and accuracy of executing morphological checking that surface under the high task demands of Experiment 3. At near-native levels, the L1 Russian speakers appear to benefit from the routines of matching case features with syntactic function in their L1, whereas their L1 English and L1 Dutch counterparts

---

7 At least for the L1 Russian advanced group which showed target offline judgements, ‘good enough’ processing is a function of task demands, while for the L1 English and L1 Dutch advanced groups, insufficient knowledge of case marking may compound difficulties with OS orders.
have difficulty at speed. In cases where the L1 and the TL use similar types of morphosyntactic information in parsing, the efficacy of processing such information in the L2 is augmented by analogous routines in L1 parsing. In a similar vein, facilitatory effects of analogous L1 morphosyntactic properties have been reported in an ERP study by, e.g., Sabourin (2003) for gender marking. Further, on-line studies on lexical retrieval in the L2 (for overview, see, e.g., Dijkstra, 2003; 2005) attest cross-linguistic influence in that the L1 is automatically activated during L2 word retrieval of cognates. Sánchez-Casas & García-Albea (2005) argue that the shared morphological representation of cognates underlies cross-linguistic activation patterns. The present findings contribute suggestive evidence that commonalities in L1-TL morphological paradigms also facilitate access to inflectional morphology in L2 parsing (for further discussion, see Chapter 9).

Seen in conjunction, these results raise the possibility of largely construing the L2 processing of morphosyntax in Experiments 1 through 3 in terms of computational load. The computational load of L2 morphosyntactic processing is determined by proficiency in the L2 and routinization of parsing and integration processes in L2 and L1 as a result of exposure and usage. Such an account of the non-native processing patterns captures the group differences according to proficiency and L1 across experiments and ties in with a number of previous findings on L2 processing. In particular, the parallels between L2ers and natives resonate with the findings on L2 inflectional morphology by McDonald (2006) and the on-line integration of information by Kilborn (1992). As in the present study, natives and L2ers evinced analogous processing and judgement patterns under different task demands, which indicates that the same grammatical and processing architecture underlies native and non-native performance, even though the extent of its accurate implementation is restricted by computational limitations of L2 processing.

6.9. **Summary**

Three experiments on the off-line knowledge and on-line processing of scrambling in L2 German showed that convergence on native processing of morphosyntax is attested at the highest proficiency levels, irrespective of L1 differences.

At the same time, the advanced groups demonstrate L1-related problems with case morphology in off-line judgements and cross-linguistic difficulties with inflectional morphology in real-time processing of the L2. For the advanced groups, these findings are in line with previous off-line and on-line research on problems with inflection (e.g. Jiang, 2004; Prévost & White, 2000b; Sabourin, 2003) as well as studies reporting flat L2 processing of non-local morphosyntactic dependencies (Marinis et al., 2005; Papadopoulou & Clahsen, 2003).

Comparisons across proficiency and L1 groups as well as across experiments indicate that non-convergence is related to a composite of proficiency and L1 effects. I suggested that these factors interact with computational load as expressed in different
Chapter 6

Task demands of Experiments 1-3 in giving rise to a systematic pattern of non-convergence at the syntax-morphology interface. This account receives support from the finding that under increased task demands in Experiments 2 and 3, the native controls evince similar selective problems as the L2ers with inflectional (case) morphology. As a consequence, I conclude that the findings from Experiment 1 through 3 are compatible with the Fundamental Identity Hypothesis, repeated in (22).

(22) Fundamental Identity Hypothesis
There are no fundamental differences between non-native and native grammatical representation or processing architecture forced by a critical period. Differences, if found, relate to factors characterizing L2 acquisition independently of a critical period, e.g. L1 transfer or performance factors, such as computational limitations, etc.

Moving beyond the morphosyntax of scrambling, Chapters 7 and 8 explore whether the Fundamental Identity Hypothesis extends to other interface aspects of scrambling in L2 German.