Online processing of syntactic constructions unique to the second language: An eye-movement study of subject–verb agreement with Swedish advanced learners of English

J.H. Greidanus

Supervisors:
Prof. Dr. M. Gullberg
Dr. R. Andersson

Abstract

It is a hotly debated issue whether second language (L2) learners can achieve target-like online syntactic processing of relations not instantiated in their first language (L1). The *Shallow Structure Hypothesis* (SSH) predicts that L2 learners will not process configurations with structural distance in a target-like online fashion (Clahsen & Felser, 2006 a,b,c), whereas some studies find that target-like processing is possible (e.g. Foucart & Frenck-Mestre, 2012). In view of the divergent findings, the present study asks how Swedish advanced learners of English (*n*=18) process complex subject–verb agreement (e.g. *the editor of the newspaper/s was/*were admired by the staff*) compared to native English speakers (*n*=15) as measured in an eye-tracking paradigm. Importantly, Swedish lacks subject-verb agreement whereas English marks it. Six conditions are investigated, in which number of the head noun (**the editor** *of the newspaper*, singular/plural), number match between head and modifying noun (*the* **editor** *of the* **newspaper**), and grammaticality (head noun/verb agreement in number) are manipulated. The results indicate that native English speakers are sensitive to manipulations of complex subject–verb agreement, with reading times varying depending on head/modifying noun match and grammaticality. In contrast, Swedish L2 learners show no difference in reading times across conditions. Moreover, more detailed analyses reveal that native speakers are more likely to skip copulas during reading than L2 readers, suggesting a greater predictive capacity and a subtle timing advantage in processing. The results are in line with the SSH prediction and previous studies of this structure (e.g. Chen et al., 2007; Jiang, 2004). Implications for theories of syntactic processing in the L2 are discussed.

Keywords: L2 processing, syntactic constructions unique to the L2, complex subject–verb agreement, eye-tracking, reading research, Shallow Structure Hypothesis, structural distance.

Table of contents

Acknowledgements

List of figures

List of tables

List of abbreviations

ANOVA	Analysis of Variance
AoA	Age of Acquisition
CEFR	Common European Framework of Reference
COCA	Corpus of Contemporary American English
(E)LAN	(Early) Left Anterior Negativity
ERP	Event-Related Potential
GJT	Grammaticality Judgement Task
L1	First Language
L2	Second Language
LexTALE	Lexical Test for Advanced Learners of English
LHQ	Language History Questionnaire
NP	Noun Phrase
PL	Plural
PP	Prepositional Phrase
QPT	Quick Placement Test
SD	Standard Deviation
SE	Standard Error
SG	Singular
SPR	Self-Paced Reading
SSH	Shallow Structure Hypothesis
TL	Target Language
TOEIC	Test of English for International Communication
UG	Universal Grammar
VP	Verb Phrase

1 Introduction

Second language (L2) processing research – a relatively young line of research – investigates how L2 learners put their knowledge of the target language (TL) to use in real time compared to native speakers (Roberts, 2013a). One of the central questions in current L2 processing research is how advanced L2 learners process grammatical structures that are not instantiated in their first language (L1) compared to native speakers of that language (Roberts & Siyanova-Chanturia, 2013). Up to this point, the findings have been rather mixed (depending on several factors to be detailed in section 2.2). The present study aims to contribute to this debate by investigating how L1 Swedish advanced learners of English process subject–verb agreement manipulations involving complex subjects online compared to native speakers of English. While the Swedish language lacks subject–verb agreement altogether, English does have a subject–verb agreement system (although a rather minimal one, see section 2.1). In this online reading study, the processing costs associated with specific subject–verb agreement manipulations are operationalised in terms of eye-movement measures. If this study does not find between-group differences for these measures, this finding will be taken to be indicative of target-like processing for the advanced learners. However, based on a review of the relevant literature, it is predicted that this study will find different processing patterns for these manipulations for Swedish advanced L2 learners of English compared to native speakers of English.

Previous research on the production of subject–verb agreement has indicated that this aspect of the TL grammar is often not fully mastered by Swedish L2 learners of English (e.g. Karlsson, 2002; Köhlmyr, 2003; Thagg Fisher, 1985). In fact, subject–verb agreement errors are among the most common errors produced by native Swedish L2 learners of English (e.g. Thagg Fisher, 1985). Källkvist and Petersson (2006) argue that this finding is not highly surprising since the acquisition of subject–verb agreement is a rather complex task, involving for example the conflation of person, number, and tense and the understanding and application of substitution (i.e. head nouns can be substituted by *he*, *she*, or *it*, which select the 3rd person form of the lexical verb in the present tense). Moreover, they state that the subject–verb agreement system in English involves several exceptions and inconsistencies, which add to its complexity (Källkvist & Petersson, 2006).

Grammatical accuracy in L2 production typically improves with proficiency. It is

hypothesised that "balanced bilinguals use procedural knowledge of syntactic and morphological rules of a language", which they apply automatically (Kormos, 2013, p. 4). In contrast, less proficient L2 learners may have procedural knowledge of specific rules, while other rules from the TL may be "stored in declarative memory and used consciously" (Kormos, 2013, p. 4). Moreover, certain syntactic rules may not have been acquired at all, in which case they may be transferred consciously from the L1 (Kormos, 2013).

Although some previous studies have focused on the production of subject–verb agreement by Swedish advanced L2 learners of English, very little is known about how this grammatical construction is processed in comprehension. Therefore, the present study asks how complex subject–verb agreement manipulations are processed in online comprehension by Swedish advanced L2 learners of English compared to native speakers of English.

2 Background

2.1 Subject verb–agreement: Swedish and English

The simple rule of subject–verb agreement (or concord) means that a singular subject selects a singular finite verb and that a plural subject takes a plural finite verb. Implementing this rule requires knowledge of the marking of grammatical number on nouns and verbs, identifying the grammatical number of a subject, and identifying the subject of a verb (Bock & Miller, 1991). As opposed to English (and many other languages), the Swedish language lacks a subject–verb agreement system. As an illustration, Table 1 shows how the copula verb 'to be' is conjugated in the simple present and past tense in Swedish and in English:

Table 1: Conjugation of 'to be' in Swedish and English

Language	Person	Pronoun	Simple present	Simple past
Swedish	1st person singular	Jag	är	var
	2nd person singular	Du	är	var
	3rd person singular	Han/Hon/Den/Det	är	var
	1st person plural	Vi	är	var
	2nd person plural	Ni	är	var
	3rd person plural	De	är	var
English	1st person singular	I	am	was
	2nd person singular	You	are	were
	3rd person singular	He/She/It	is	was
	1st person plural	We	are	were
	2nd person plural	You	are	were
	3rd person plural	They	are	were

Table 1 indicates that, in Swedish, the verb is conjugated for tense, but not for person or number. The systematic difference between Swedish and English can be seen in example (1):

(1)

    a. *Swedish*: [$_{NP}$ **redaktören**$_{SG}$ [$_{PP}$ på [$_{NP}$ tidningen]]] **var_** beundrad av personalen (…)

        English: [$_{NP}$ the **editor**$_{SG}$ [$_{PP}$ of [$_{NP}$ the newspaper]]] **was**$_{SG}$ admired by the staff (…)

    b. *Swedish*: [$_{NP}$ **redaktörerna**$_{PL}$ [$_{PP}$ på [$_{NP}$ tidningen]]] **var_** beundrade av personalen (…)

        English: [$_{NP}$ the **editors**$_{PL}$ [$_{PP}$ of [$_{NP}$ the newspaper]]] **were**$_{PL}$ admired by the staff (…)

In the Swedish version of this example, the copula verb remains *var* when moving from the singular to the plural version of the complex subject's head noun, from *editor* to *editors*. In English, on the other hand, the subject–verb agreement relation entails that "a finite verb agrees

in number with the highest noun in a hierarchical phrase-structure-representation of the clause" (Eberhard, 1999, p. 560). In the English version of example (1), this is reflected in the fact that the past-tense copula changes in number from *was* to *were*, in accord with the number change of the complex subject's head noun.

Although the above example provides a good illustration of how subject–verb agreement may function in English, the agreement situation in this sentence construction is rather unique. That is, in English only the copula verb 'to be' is marked in first and in third-person and in present- as well as in past-tense forms (Bock & Miller, 1991; Pearlmutter, Garnsey, & Bock, 1999). Other finite verbs are only marked for number in third-person present-tense form (simple present and present perfect); there is no number marking for other past-tense verbs or verb constructions involving any modal. In this sense, the subject–verb agreement system in English has rather limited applicability and reliability (MacWhinney, Bates, & Kliegl, 1984).

Moreover, whereas the computation of subject–verb agreement in English is a prerequisite for language producers to be able to construct a broad spectrum of grammatical sentences, it is often not obviously required from the point of view of sentence comprehension. In their 1984 study, MacWhinney et al. presented English, German, and Italian subjects with simple transitive sentences in which word order, agreement, animacy, and stress contrasts were varied. Subjects were asked to decide which noun – out of two nouns in each of the experimental sentences – was the actor. They found that Italian and German subjects relied heavily on agreement cues (as well as on animacy, in the latter case), whereas native English participants relied strongly on word order cues. Therefore, they concluded that morphological cues are "an unreliable source of information about sentence relations in English", whereas word order is "reliable and omnipresent" (MacWhinney et al., 1984, p. 135). Pearlmutter et al. (1999) also emphasised that the English word order is relatively fixed, "with the result that the subject NP and the main verb of a clause are nearly always identifiable on the basis of positional and syntactic category information alone" (p. 428). However, despite the relatively fixed word order and the meagreness of the English agreement system, the latter comprehension study found that participants displayed early sensitivity to agreement violations, indicating that grammatical number features were in fact accessed and processed continuously.

2.1.1 Agreement attraction

Much previous work on the production of complex subject–verb agreement (e.g. Bock & Miller, 1991; Eberhard, 1997; Eberhard, 1999; Eberhard, Cutting, & Bock, 2005; Franck, Vigliocco, & Nicol, 2002; Solomon & Pearlmutter, 2004) has demonstrated so-called *attraction errors* (Wagers, Lau, & Phillips, 2009). This type of error has been found to occur when the grammatical number of an intervening local noun, an 'attractor' noun, does not match the grammatical number of the complex subject's head noun, resulting in erroneous agreement between the verb and the local noun (Levy, 2013), as in example (2). In this example, the past-tense copula erroneously agrees with the local noun in number:

(2)　* The editor$_{SG}$ of the **newspapers$_{PL}$ were$_{PL}$** admired by the staff (…)

It should be noted that this type of error in production typically occurs when the head noun is singular, not when it is plural (i.e. explicitly marked for number). Within the sentence production literature, there are roughly two leading sets of theories explaining this agreement attraction phenomenon. The first class of theories suggests "that these effects are rooted in faulty representation of the number of the subject" (Wagers et al. 2009, p. 206). These theories are sometimes referred to as 'percolation theories', since they predict that "attraction effects arise from percolation of agreement features from within a complex NP", "up to the NP head, leading to incorrect representation of the subject NP's number" (Levy, 2013, p. 86). Applied to example (2), this would mean that the plural number feature from the local noun *newspapers* would percolate up to the singular head noun *editor*.

The second class of theories rather suggests "that such effects arise in the process of re-accessing subject number at the verb" (Wagers et al. 2009, p. 206). That is, features of simultaneously activated constituents (in this case the grammatical number features of the head noun and the local noun) may be confused when information about the complex subject is re-accessed at the verb.

A few comprehension studies have also been conducted. In their 1999 comprehension study, Pearlmutter et al. found that singular verb versions (3)a and (3)b were comparatively easier to process than plural verb versions (3)c and (3)d, but also that plural attractor nouns – *offices* in

the example below – reduced the processing costs for ungrammatical conditions (as reflected in shorter reading times immediately after the verb).

(3)
    a.   The door to the office was (…)

    b.   The door to the offices was (…)

    c.   *The door to the office were (…)

    d.   *The door to the offices were (…)

This finding is explained rather satisfactorily by the percolation theories. That is, if the number of the complex subject is misinterpreted, the agreement anomaly might not be recognised, as reflected in lower processing costs. However, Wagers et al. (2009) presented contradicting evidence. In their experiments, they varied the structural distance between the subject noun and the attractor noun. They found that the attraction effect also occurs in relative clause configurations where the attractor noun – *musicians* in (4)b – does not directly intervene between the subject and the verb. That is, the plural number marking on the noun positioned outside of the relative clause (i.e. in a hierarchically superior position) was found to reduce or even eliminate the processing cost for the anomaly, even though the verb's subject is the relative-clause internal and singular NP *the reviewer*.

(4)
    a.   *The musician who the reviewer praise so highly will probably win a Grammy.

    b.   *The musicians who the reviewer praise so highly will probably win a Grammy.

Since the plural noun is not embedded in the subject of the relative clause, upward percolation of its plural number feature would not lead to a misrepresentation of the subject's number. Furthermore, in six self-paced reading (SPR) experiments where participants were asked to do speeded sentence acceptability judgements, Wagers et al. (2009) found that the attraction effect only occured for ungrammatical constructions. This finding is unexpected from the point of view of the percolation theories. That is, if the attraction effect would indeed reflect errors in subject number encoding, it would logically follow that in grammatical constructions with a singular head noun and a plural attractor noun, the singular verb would in some cases be perceived as ungrammatical. This was not observed from the total reading times and the acceptability rating scores in the said study. Therefore, Wagers et al. argued that "agreement attraction in

comprehension arises in the process of re-accessing information about the subject on the basis of cues at the verb, and not because the number of the subject is incorrectly represented" (2009, p. 207).

2.1.2   Conceptual (notional) number of nouns

Another relevant concept is the conceptual (or notional) number of nouns. The conceptual number of a noun expresses "whether its referent is one entity or multiple entities" (Eberhard, 1999, p. 561). The grammatical number of a noun, on the other hand, corresponds to its singularity or plurality, which is often but not always reflected in its morphological form (Eberhard, 1999). For some categories of nouns, conceptual and grammatical plurality are found to diverge (Bock & Miller, 1991). Examples are nouns that are conceptually singular but grammatically plural and that take plural number marking on the verb (*scissors, jeans*), as well as nouns that are conceptually plural but grammatically singular and that take plural number marking on the verb (*police, people*).

An example of a complex subject in which the conceptual number and grammatical number diverge is *the picture on the covers*. This subject is singular in grammatical number, but more ambiguous in conceptual number. That is, from this example we understand that each cover has a picture; therefore there are multiple pictures (even if it is always the same one). Several studies, mostly of monolingual, but in some cases of bilingual speakers (e.g. Bock & Miller, 1991; Eberhard, 1999; Hoshino, Dussias, & Kroll, 2010; Nicol & Greth, 2003; Pearlmutter et al., 1999), have investigated the effects of mismatches in grammatical and conceptual number on the production (and comprehension) of (complex) subject–verb agreement. The findings were not particularly straightforward even in the monolingual literature (i.e. in most cases, no clear effects of mismatches between conceptual and grammatical number were found). Therefore, the stimuli in the present study will not be balanced for conceptual number, nor will grammatical and conceptual number (mis)matches be considered as a possible predictor.

2.2   Concepts and theories in L2 syntactic processing research

It has been found that advanced L2 learners' syntactic production and comprehension may be asymmetric, i.e. even when production is persistently non-target-like, (online) comprehension may still be target-like (Trenkic, Mirkovic, & Altmann, 2013). Moreover, there may be

asymmetry between online and offline comprehension. For example, Tokowicz and MacWhinney (2005), using the event-related potential (ERP) technique and a grammaticality judgement (GJT) paradigm, investigated whether English L2 learners of Spanish showed online and offline sensitivity to violations of grammatical structures in the TL. Interestingly, they found online sensitivity to violations of specific structures for which the grammatical judgement accuracy was near chance. Taken together, these results imply that offline sensitivity is not a prerequisite for online sensitivity. Likewise, it is well-attested that offline sensitivity (as tested in e.g. grammaticality judgement tasks) does not imply online sensitivity (e.g. Hopp, 2010; Jiang, 2004; Ojima, Nakata, & Kakigi, 2005).

A major debate in the field of L2 processing research revolves around the question whether online syntactic processing of the TL is qualitatively or fundamentally different for advanced L2 learners compared to native speakers (Roberts, 2013b). Roberts (2013b) states that "it is uncontroversial that native speakers incrementally process the language input, making rapid use of bottom–up lexical-semantic and syntactic information as well as top–down, discourse-pragmatic information as new material is integrated into the parse during real-time processing" (p. 221). Studies in L2 processing research aim to investigate to what extent these L1 processing characteristics apply to L2 processing as well, and to what extent differences that are found between L1 and L2 learner processing are qualitative or rather quantitative in nature.

The concepts of qualitative versus quantitative differences between native speaker processing and L2 learner processing need to be fleshed out further in terms of methodologies that are commonly applied in psycholinguistic research, for example ERP, self-paced reading, and eye-tracking (during reading). Although qualitative and quantitative differences cannot always be teased apart easily (i.e. they are typically connected), assuming fairly homogeneous and reasonably large participant groups, qualitative between-group differences are characterised more absolutely by the presence versus absence of certain effects/components, or by a difference in their direction (and/or their distribution; in ERP studies, an effect may for example be more localised for native speakers). For example, in an ERP study involving a native speaker group and an L2 learner group, a specific ungrammatical condition may result in the presence of a clear component for the native speaker group (compared to a baseline), whereas no such component (e.g. as in Ojima et al., 2005) or a component in the opposite direction (e.g. as in Chen, Shu, Liu, Zao, & Li,

2007; Foucart & Frenck-Mestre, 2012) may be found for the learner group. Likewise, in SPR studies or eye-tracking studies, a certain condition may significantly inflate reading times for a specific critical region for native speakers (compared to a baseline condition), whereas no statistically significant effect (e.g. Jiang, 2004; Keating, 2009) or potentially an effect in the opposite direction (i.e. significantly deflated reading times) may be found for L2 learners.

Quantitative differences, on the other hand, are instead reflected in the actual strength and/or latency of effects/components. For example, in an ERP study involving an L2 learner group, between-group amplitude differences may be found for a component following from a certain manipulation, i.e. a specific component may be visible for both groups, but its amplitude may be lower for the learner group (e.g. Foucart & Frenck-Mestre, 2012). Likewise, in SPR studies and in eye-tracking studies, reading times may differ for native speakers and learners. Importantly, reading times are typically longer for learners, for example because of slower lexical access and parsing (which may still be qualitatively similar). To be able to distinguish between qualitative and quantitative differences, it is always important to compare effects for specific manipulations between groups as well as within groups (i.e. taking into account the groups' own baselines).

Furthermore, quantitative differences may be reflected in latency differences. That is, an ERP component for L2 learners may be observed comparatively late in a specified time window compared to that for native speakers (e.g. Hahne, 2001; Sabourin & Stowe, 2008). Similarly, in eye-tracking reading studies, effects for learners may be delayed, as reflected in sensitivity only in late eye-tracking measures (see section 4.2.2) for L2 learners versus sensitivity both in early and in late measures for native speakers (e.g. Keating, 2010). Moreover, both in eye-tracking- and SPR-paradigms, an effect may appear more immediately for native speakers, while it only appears on a later word (or words) for L2 learners. To clarify, late sensitivity implies "a quantitative difference in processing, not a qualitative one", because L2 learners in fact show an ability to detect anomalies online, albeit "not as early as native speakers" (Keating, 2010, p. 128).

One set of theories, culminated in for example the Shallow Structure Hypothesis (SSH) proposed by Clahsen and Felser (2006 a,b,c), argues that syntactic processing is qualitatively different for adult L2 learners compared to native speakers of the TL. The SSH states that L2 learners rely more heavily on lexical, pragmatic, and contextual cues rather than on syntactic ones

(Clahsen & Felser, 2006b) and that "the syntactic representations adult L2 learners compute for comprehension are shallower and less detailed than those of native speakers" (p. 32). This means that target-like processing of morphosyntactic features is limited to local domains, which are rather loosely defined as "closely adjacent constituents" (Clahsen & Felser, 2006a). That is, whereas the SSH predicts that for example gender agreement computation within a phrase may be target-like (qualitatively similar), processing of structural dependencies across phrases or clauses is predicted to be qualitatively different. The prediction of the SSH for simple subject verb-agreement (e.g. *the lawyer was/*were*), involving a local dependency, is that target-like processing may occur under certain circumstances (e.g. Ojima et al. (2005) found a target-like component for the advanced learner group in their ERP study). However, this may not apply to complex subject–verb agreement (e.g. *the lawyer on the case was/*were*) – the type of agreement investigated in the present study. The complex subject as a whole receives its grammatical number marking from its head noun, which agrees in number with the finite verb. In the materials used in the current study, a prepositional phrase (PP) creates structural distance between the features marked for grammatical number. Therefore, complex subject–verb agreement should be seen as a non-local dependency. According to the SSH, manipulations of this type of structure will not be processed in target-like ways by advanced learners of the TL.

Other theories assume that L2 learners do have access to similar parsing mechanisms available for monolingual processing but that L2 learners oftentimes perform differently because of for example "insufficient L2 proficiency and/or cognitive capacity limitations, like speed and working memory" (Roberts, 2013b, p. 221). These accounts assume quantitative rather than qualitative processing differences between native speakers and advanced learners of a given language (e.g. Foucart & Frenck-Mestre, 2012; Hopp, 2010; Sabourin & Stowe, 2008). The present study mainly aims to establish to what extent complex subject–verb agreement is processed in qualitatively similar or different ways by advanced L2 learners compared to native speakers.

2.2.1 Factors affecting L2 syntactic processing

The age at which the acquisition of the L2 starts (the age of acquisition, henceforth AoA) has traditionally been considered as the main factor predicting the level of ultimate attainment of L2 proficiency (Abrahamsson & Hyltenstam, 2008). The ability to acquire an L2 to native-like levels

(in offline as well as in online terms) typically decreases with increasing age, but may be mitigated by factors such as time, exposure, motivation, and aptitude (i.e. the more time, motivation, and aptitude for language learning and the more exposure to the L2, the more likely an L2 learner is to attain native-like proficiency) (Abrahamsson & Hyltenstam, 2008). The correlation between AoA and ultimate attainment is an important point of discussion in the debate surrounding the existence of a *critical period* for aspects of L2 learning and the existence of and access to *Universal Grammar* (UG) (for a comprehensive overview of UG theory, see e.g. Cook & Newson, 2007). It should be noted that critical period approaches do not assume that it is impossible to successfully learn an L2 after a certain age, but rather that there may be qualitative differences, especially for aspects of morphosyntax (e.g. Weber-Fox & Neville, 1996). Lexical-semantic learning and processing, on the other hand, are typically more robust and more independent of AoA (e.g. Ojima et al., 2005).

Proponents of a critical period have for example suggested the *Failed Functional Features Hypothesis* (Hawkins & Chan, 1997) that assumes that persistent problems with L2 morphosyntax are best characterised as a representational deficit. The hypothesis states that a "divergence from native-speaker representations is an effect of the inaccessibility of features of functional categories in second language acquisition" (Hawkins & Chan, 1997, p. 187). Opponents have for example advocated the *Full Transfer / Full Access Hypothesis* (Schwartz & Sprouse, 1994, 1996). This hypothesis says that all "principles and parameter values as instantiated in the L1 grammar immediately carry over as the initial state of a new grammatical system on first exposure to input from the target language" (Schwartz & Sprouse, 1996, p. 41) and that UG is fully accessible for restructuring the incoming input from the TL (Schwartz & Sprouse, 1996). Approaches such as these predict that native-like attainment should be possible and that persistent problems are for example caused by limitations in L2 processing efficiency.

Against approaches supporting a critical period hypothesis for L2 learning, several recent online studies have indicated that qualitatively similar syntactic processing may be possible for late learners under certain circumstances, hence that other factors may be more important than AoA (e.g. Foucart & Frenck-Mestre, 2012; Gillon-Dowens, Guo, Guo, Barber, & Carreiras, 2011; Gillon-Dowens, Vergara, Barber, & Carreiras, 2010; Hahne, 2001; Hahne, Mueller, & Clahsen, 2006; Hopp, 2010; Keating, 2009; Roberts & Liszka, 2013; Rossi, Gugler, Friederici,

& Hahne, 2006; Sabourin & Stowe, 2008; Steinhauer, White, King, Cornell, Genesee, & White, 2006). The factors considered and investigated in these studies typically include the L2 proficiency level, as well L1–L2 typological similarities and differences, the possibility of L1 transfer, linear distance (number of intervening words) and structural distance (number of intervening phrases) in the linguistic configurations under scrutiny, and, to a lesser extent, factors such as working memory capacity, attention, and speed.

It is well-attested that L2 proficiency is a strong predictor of the target-likeness of L2 syntactic parsing and it is often varied or heavily controlled for in L2 syntactic processing studies. For example Keating (2009) and Hopp (2010) only found target-like processing for violations of specific grammatical structures provided that their learners had reached advanced or near-native L2 proficiency levels, respectively. Factors such as working memory are often controlled for rather than investigated explicitly (Roberts, 2013b). However, Service, Simola, Metsänheimo, and Maury (2002) found some evidence that certain aspects of comprehension in the L2 may require less working memory resources when L2 learners become more experienced/proficient.

In general, "there is a marked lack of L1 influence on bilinguals' grammatical processing, and so one cannot always attribute L1–L2 processing differences to this factor" (Roberts, 2013b). Several studies have indicated that learners from different L1-backgrounds often pattern together and differently from natives – even if the L1 and the L2 are highly similar (e.g. Marinis, Roberts, Felser, & Clahsen, 2005; Roberts, Gullberg, & Indefrey, 2008). Another interesting finding is that structures that are formed similarly in the L1 and L2 as well as structures that are absent in the L1 are typically more likely to be processed in target-like ways compared to configurations that function differently in the L1 and L2 (e.g. Tokowicz & MacWhinney, 2005). Tokowicz and MacWhinney (2005) found that, while English learners of Spanish showed online sensitivity to violations of configurations that were either similar or absent in their L1, they were not sensitive to violations of structures that were constructed differently in their L1. The authors explained these findings according to the principles of the *Competition Model* (MacWhinney, 1987, 2005). They suggested that positive transfer and a lack of online competition explain why structures that function similarly in the L1 and L2 may be processed in target-like ways. Structures that are computed differently in the L1 and L2, on the other hand, are subject to negative transfer and online competition, which may prevent target-like processing (cf. Ellis, 2006). For structures

unique to the L2, the model predicts that target-like processing may be possible, since these are not affected by negative L1 transfer or by online competition.

2.2.2 Previous research on online processing of agreement relations

Several studies have investigated the online processing of different types of agreement relations in the L2, with rather mixed results. This section will present relevant findings, mainly from studies on the online processing of agreement relations unique to the L2. Although the present study revolves around subject–verb agreement, other types of agreement will be discussed in this overview.

Jiang (2004) used the SPR paradigm to investigate how Chinese advanced learners of English processed complex subject–verb agreement manipulations online compared to native speakers of English. In three experiments, four out of six conditions from Pearlmutter, et al. (1999) were adopted, alongside person–verb agreement and verb subcategorisation manipulations. The study found qualitatively different processing patterns for subject–verb agreement violations for the learner group compared to the native speaker group. That is, learners did not show significant effects for ungrammaticalities or local mismatches, as opposed to native speakers. The other idiosyncrasies tested did reveal some online sensitivity in the learner group. Jiang (2004) argued that these findings indicate that morphological knowledge is not internalised in these learners, while the computation of agreement (person–verb) is.

The conclusions that were drawn in this study do not seem fully justified, however. Jiang (2004) stated that distance (it is not made explicit whether linear or structural distance is meant) cannot account for the learners' insensitivity to subject–verb agreement manipulations, since learners were found to be sensitive to violations in non-local verb categorisation conditions. However, it should be noted the sensitivity in the learner group was only marginally significant (in the by-subjects analysis), in one region of interest (compared to by-subjects and by-items significance in two regions of interest for the native speaker group). Furthermore, verb subcategorisation is rather lexically driven, and non-local subcategorisation violations might therefore not be directly comparable to non-local agreement violations (more semantically driven processing is often found to be more robust for L2 learners).

Wen (2007), in a similar SPR study, investigated how Japanese learners of English processed non-local subject–verb agreement manipulations, involving regular verbs (e.g. *attract*)

instead of the copula *to be*, compared to native speakers of English. In line with Jiang (2004), the study found qualitatively different processing patterns for learners compared to native speakers, implying that the learners were not sensitive to violations of subject–verb agreement. Jiang (2007), in another SPR study, involving number violations within complex noun phrases (*several of the rare coins/*coin*), replicated the finding that learners with an L1 background that does not encode grammatical number may not be able to process number marking in L2 English in target-like ways. It should be noted that the learners in Jiang (2004, 2007) and Wen (2007) were all from an L1 background (Chinese or Japanse) that is very remote from English, and Jiang (2004) himself suggested that testing learners from different (and less remote) L1 backgrounds may provide more conclusive answers. Importantly, grammatical number (and consequently number agreement) was not instantiated in the learners' L1 and this may have caused non-target-like processing of number violations. However, against this assumption, Wen, Miyao, Takeda, Chu, and Schwartz (2010), in an SPR study with Chinese and Japanese intermediate and advanced L2 learners of English, found online sensitivity to violations of demonstrative-noun number agreement (*this beautiful house/*houses*) in the advanced learner groups. Importantly, the materials in the said study involved a linear (and not a structural) distance, created by the presence of an adjective between the (dis)agreeing elements. Since this study found online sensitivity despite the linear distance involved, the authors argued that strict adjacency is not a prerequisite for target-like processing. They also cautiously suggested that structural rather than linear distance may have been at the root of non-target-like processing found in earlier studies (Wen et al., 2010). This study also found clear proficiency effects, with only the advanced learners showing target-like parsing.

Hopp (2010) investigated language pairings that were less remote. This study employed the SPR paradigm to investigate whether L1 Dutch, English, and Russian advanced to near-native L2 learners of German were able to process non-local subject–verb agreement disambiguation offline and online in qualitatively similar ways as native speakers of German. Subject–verb agreement was instantiated in all L1s of the learner groups involved. The study found that near-native speakers processed subject–verb agreement incrementally (i.e. the natives' and near-natives' processing patterns were virtually indistinguishable), whereas the advanced learners did not show target-like processing. The results of the speeded offline grammaticality

judgements indicated that the advanced learner groups were affected by L1 transfer, whereas no L1-effects were found for any of the near-native groups. Taken together, the author took these results as evidence that proficiency is a more important predictor than AoA. He also suggested that any effect of L1 transfer may be overcome at near-native proficiency levels.

Two ERP studies have also examined subject–verb agreement (Chen et al., 2007; Ojima et al., 2005). The ERP components of interest for studies on syntactic parsing are (early) left anterior negativity ((E)LAN) and the P600 that are typically observed to follow from morphosyntatic violations in monolingual studies (e.g. Friederici, 2002; Ingram, 2007). Compared to the (E)LAN, the P600 component is typically more stable (Frenck-Mestre, 2005). Ojima et al (2005) looked at how Japanese intermediate and advanced L2 learners of English processed violations in simple subject–verb agreement constructions, such as (5):

(5) Plants grow/*grows from seeds

The study found a left-lateralised negativity resembling left anterior negativity for the native group and the advanced learner group, but this component was absent for the intermediate learner group. Additionally, a P600 component was found for the native speaker group, but it was absent for both learner groups. Importantly, semantic violations investigated in the same study evoked a clear N400 component in the native speaker group as well as both learner groups (with a processing delay for the learner groups, however). The authors argued that semantic processing is more stable and robust compared to syntactic processing, which may be subject to critical-period effects.

Chen et al. (2007) investigated how Chinese L2 learners of English processed complex subject–verb agreement violations of the type adopted in the present study compared to native speakers of English. The constructions involved a singular head noun, followed by a PP and a verb phrase (VP). In the PP, the grammatical number of the noun was varied and in the VP, the number of the copula verb was varied, as in (6):

(6) The price of the car(s) was/were too high

The study thus involved two grammatical conditions and two ungrammatical ones. Whereas the offline data from a grammaticality judgement task showed that the learners were sensitive to the violations offline, the ERP responses did not show a typical biphasic (LAN/P600) pattern for the

learners, as opposed to the native speakers. Interestingly, the ungrammatical conditions revealed a negative peak in the window in which the P600 (a positive peak) typically occurs for native speakers. The grammatical incongruent condition (*singular, plural, singular*) elicited an N400 (associated with semantic processing difficulties) followed by a P600 for the learners, which the authors attributed respectively to the semantic or conceptual incompatibility of the two nouns and the incongruence in the number of the local noun and the copula verb. These results seem to converge with the SSH predictions that target-like processing is not possible for non-local domains, as opposed to local domains.

Two eye-tracking studies that investigated gender agreement phenomena are Keating (2009) and Foucart & Frenck-Mestre (2012). Keating (2009) investigated how groups of English L2 learners of Spanish with different proficiency levels (beginner, intermediate, and advanced) processed violations of grammatical gender in the TL (importantly, English lacks grammatical gender). In this study, the distance between the head noun and the adjectives violated for gender was varied (the violations were within-phrase, across-phrase, or outside the matrix clause), specifically to test the predictions of the SSH (Clahsen & Felser, 2006 a,b,c). The results indicated that within-phrase violations were processed in similar ways by native speakers and English advanced L2 learners of Spanish, whereas across-phrase violations and gender violations outside the matrix clause were processed differently by native speakers compared to all other groups. These findings were in line with the predictions of the SSH. However, it was not clear whether this distance effect was a linear or a structural phenomenon and to what extent "individual differences in working memory capacity" affected "learners' sensitivity to gender errors" (Keating, 2009, p. 529). In a 2010 follow-up study, Keating investigated these issues and found that linear distance, keeping structural distance constant, affected both native speakers' and advanced L2 learners' sensitivity to gender anomalies (although at different points in the eye-movement record). Moreover, individual differences in working memory were shown to have an effect in the advanced learner group.

Foucart and Frenck-Mestre (2012), in a study based on both ERP and eye-tracking methodologies, investigated grammatical gender processing in English advanced learners of French. In contrast to Keating (2009) and against the SSH predictions, the eye-tracking data

suggested "early and equivalent detection of gender agreement violations during online sentence processing for native and non-native speakers" (p. 244), even in non-local conditions, as in (7):

(7) Au printemps, les pommes_{fem} sont vertes_{fem}/*verts_{masc} sur cet arbre

   In the spring, the apples are green on this tree

However, whereas violations of all structures consistently evoked P600 components in the native speaker group, the learner group showed a P600, an N400, or no effect depending on the structure that was violated. Thus, an asymmetry between eye-tracking data and ERP data was found for violations of non-canonical and non-local dependencies (see also section 2.3.). From the eye-tracking results, it was concluded that grammatical gender processing may be qualitatively and quantitatively similar for learners and native speakers. The ERP results, on the other hand, indicated both qualitative (i.e. a component in the opposite direction or no component) and quantitative (lower amplitude) differences between the groups. Noticeably, this study did not find processing delays for the advanced learner group.

   Alemán Bañón, Fiorentino, and Gabriele (2014) employed the ERP paradigm to investigate how native English L2 learners of Spanish processed grammatical number- and gender violations in the TL. Whereas grammatical number is encoded in English, the language does not instantiate grammatical gender. Therefore, a comparison of the ERP responses for the two structures was expected to provide insights into whether or not target-like processing is possible for structures unique to the L2. Moreover, there are differences in how Spanish and English realise number agreement. Whereas both Spanish and English instantiate demonstrative–noun agreement, only Spanish instantiates adjective–noun agreement. The authors were interested in finding whether both structures could be processed in target-like ways by L2 learners. Finally, they varied the structural distance (the linear distance was kept constant) between (dis)agreeing features – within-phrase and across-phrase – to test the predictions of the SSH. Their results indicated that target-like processing of all structures was possible, in line with the Full Transfer / Full Access Hypothesis (Schwartz & Sprouse, 1996). The size of the components was found to be larger for number violations than for gender violations, which applied to both the learner group and the native speaker group (cf. Gillon-Dowens et al., 2010). Structural distance affected the size of the components both for learners and for native speakers, but the components were present both in local and in non-local conditions, contra the SSH.

To summarise briefly, processing of agreement relations may be affected by learners' proficiency, linear and structural distance in the structures under scrutiny, L1–L2 typological differences and similarities, and L1 transfer. There is inconsistency in the results, however, with some studies indicating that target-like processing is possible when the agreement structure (or even grammatical number in general) is not instantiated in the learners' L1, irrespective of structural distance between the agreeing features, and others presenting opposing evidence.

The present study aims to resolve some important issues with the studies discussed above. First of all, this study will control more tightly for word-level factors (see sections 4.2.1 and 4.3.2) compared to several of the studies discussed above. In most of these studies, there is no or hardly any mention of word-level factors guiding the design of the stimuli, whereas for example lexical frequency and contextual plausibility (see section 4.2.1) are relevant, particularly for studies involving learner groups. Secondly, eye-tracking will be adopted as a method. In section 4.2, it will be argued that this method is highly suitable for investigating syntactic parsing online in learner groups, especially when non-local dependencies are involved. That is, the eye-tracking method places less of a burden on the working memory compared to SPR and ERP paradigms. Finally, the study will investigate a more closely related language pairing, in which both languages overtly encode grammatical number. Although Swedish does not instantiate subject–verb agreement, other types of number agreement – such as determiner–noun and adjective–noun agreement – are instantiated. Therefore, it may be expected that L1 Swedish L2 English learners are potentially more sensitive to violations of number agreement in English compared to learners from more remote L1 backgrounds that lack any form of number agreement (such as Japanese and Chinese). Importantly, the present study aims to control for the possibility that non-target-like processing of complex subject–verb agreement by L2 learners is either method-induced or caused by considerable L1–L2 typological differences. If non-target like processing patterns are found for the learners (as is predicted), this finding can be attributed to factors such as structural distance and structural complexity (and possibly limitations in proficiency) rather than to methodological issues or the remoteness of the L1. If the processing is found to be target-like, however, this will be taken as evidence against the SSH and other structural difference accounts.

3 The current study

3.1 Aims and hypotheses

This study aims to examine whether advanced L2 learners can process grammatical structures that are not present in their L1 in qualitatively similar ways as native speakers. We investigate the SSH prediction that target-like processing is not possible for non-local domains (Clahsen & Felser, 2006 a,b,c) by employing a configuration involving a structural distance between the agreeing elements. To this end, eye-movement patterns of Swedish advanced L2 learners of English will be compared to those of native speakers of British English. That is, we will investigate whether native speakers and advanced learners respond similarly to different manipulations, for example as revealed by statistically significantly longer reading times for ungrammatical versus grammatical conditions.

The structure under scrutiny is subject–verb agreement, instantiated in English, but not in Swedish. The present study builds largely on Pearlmutter et al. (1999). In line with this monolingual study, the experimental sentences appear in six conditions, two of which are ungrammatical (see section 4.3.1). Within the complex subject in the experimental sentences, the grammatical number of the head noun and the number of the local noun are varied. Additionally, the number of the copula verb is varied, resulting in six different conditions, see example (9), p. 30.

3.2 Predictions for the native speakers

In line with the findings from Pearlmutter et al. (1999, condition a (the *baseline*; the *editor* of the *newspaper was*), in which the head noun, the local noun, and the copula verb are all singular, is expected to be easier to process compared to the other conditions. That is, in this condition there is no explicit number marking that needs to be parsed and retained in working memory. Additionally, in many cases this condition is semantically most plausible. If the native speakers in the present study match the those in Pearlmutter et al. (1999), the first-pass reading times and the total reading times for this condition will generally be shorter than those for other conditions (see section 4.2.2 for definitions of these eye-movement measures).

The other conditions, and especially the two ungrammatical conditions, are expected to induce comparatively heavier processing costs, as reflected in longer first-pass reading times,

longer total reading times, and higher first-pass regression probabilities. However, echoing the findings from Pearlmutter et al. (1999), plural attractor nouns are expected to somewhat deflate the first pass-reading times and the total reading times. Therefore, condition d (*the *editor* of the *newspapers were*) is expected to incur lower processing costs compared to condition c (*the *editor* of the *newspaper were*). The probability of a first-pass regression from the 'verb region' *was admired* for condition c is expected to be higher than that for any of the other conditions. Interestingly, Pearlmutter et al. (1999) found that condition e (the *editors* of the *newspapers were*) in which the plural grammatical number of both nouns and the verb align, comes with heavier processing costs compared to condition f (the *editors* of the *newspaper were*), in which the local noun is singular. The present study is expected to replicate these findings for the native speakers.

3.3 Predictions for the advanced learners

Based on a review of relevant literature (2.2.2.), the present study expects to find non-target-like eye-movement patterns for the advanced learners. While the linear and structural distance between the (dis)agreeing features in studies such as Foucart and Frenck-Mestre (2012) and Alemán Bañón et al. (2014) did not prevent target-like processing, the linear and structural distance in the present study are somewhat larger and may prevent target-like processing. Moreover, these studies investigated a different type of agreement (i.e. adjective–noun) and the structures tested were apparently less structurally complex (i.e. in both studies, a copula verb interfered between the (dis)agreeing elements, while a more complex PP, containing an attractor noun, creates distance in the items used in the present study).

More specifically, the first condition (a), in which both nouns and the copula verb are singular, is expected to induce lower processing costs compared to the other conditions (as is also predicted for the native speaker group). Both for native speakers and for L2 learners, the first-pass reading times and total reading times are expected to be inflated somewhat for the other three grammatical conditions (b, e, f), since the plural marker on either or both of the nouns needs to be processed and retained in working memory for participants to be able to comprehend the sentences in full. In case of a plural local noun (b, e), there may be a slight spillover effect (on the verb region). The copula verb in the two ungrammatical conditions and in the two plural head noun grammatical conditions (c, d, e, and f) is also one character longer compared to the copula in conditions a and b, which may inflate fixation durations in this region somewhat. Apart from

these syntactic and physical (length) differences, there are also semantic differences between conditions that could cause variability in reading times. However, this is strongly dependent on the experimental sentence in question and therefore less predictable. That is, in some cases the b-condition (the *editor* of the *newspapers was)* may for example be less semantically plausible than the f-condition (the *editors* of the *newspaper were*), while it may be the other way around in others.

For the L2 learners, the two ungrammatical conditions (c and d) are expected to show an effect reversed to that predicted for the native speakers. That is, a plural local noun is expected to inflate the fixation durations somewhat for the L2 learners (the processing cost induced by the plural marker on the local noun may spill over onto the next region to some extent), whereas for the native speakers it is expected to deflate the fixation durations because of the presence of an attraction effect. The probability of a first-pass regression from the verb region is expected to be rather similar for all conditions, although possibly somewhat larger for the ungrammatical conditions.

4　Method

4.1　Participants

Participants were recruited via social media or via e-mail. Two so-called public 'event pages' were created on the social media platform Facebook. British and Swedish participants could find the inclusion criteria and important details about the experiment (dates, location, duration, compensation etc.) in the event descriptions, after which they were free to decide whether they were willing to participate. In case they signed up for the experiment, they were contacted by the experiment leader for an appointment. A number of participants were contacted because they appeared on lists of people who had previously participated in (eye-tracking) experiments at Lund University Humanities Lab. All participants reported normal or corrected to normal vision. Participants were compensated for their participation with a cinema ticket (or alternatively with SEK 100) and sweets.

4.1.1　Native speaker group

The native speaker group consisted of 18 native speakers of British English (8 male), mean age 22;11 (22 years; 11 months), range 20–32. Native English speakers came from England, Scotland, Northern Ireland, or Wales. Functional bilinguals (including English-Irish and English-Welsh bilinguals) could not participate. Moreover, native speakers of British English with self-reported knowledge of any other L2 (especially Swedish) above B1/B2-level in terms of the Common European Framework of Reference (CEFR) levels (see section 4.3.4) could not participate. Furthermore, participants could not be enrolled in a languages or linguistics programme, and they could not have been diagnosed with dyslexia. One British participant was excluded from analyses due to a failure to comply with important inclusion criteria, another because of low reading comprehension accuracy (57%), and a third participant because of poor data quality. This led to a native speaker group of 15 native speakers of British English (7 male), mean age 22;6, range 20–27.

4.1.2　Advanced learner group

The advanced learner group consisted of 22 native Swedish advanced learners of English (18 male), mean age 24;2, range 20–30. Swedish native speakers were required to have an advanced

(self-reported) proficiency level in English (C1-level according to the CEFR), and they could not speak any language other than English at an advanced level. Furthermore, participants could not have lived in an English-speaking country for more than 6 months. Finally, participants could not be enrolled in a languages or linguistics programme and they could not have been diagnosed with dyslexia. Two Swedish participants were excluded from analyses because they failed to comply with important inclusion criteria and two Swedish participants were excluded due to poor data quality. This led to an advanced learner group of 18 participants (15 male), mean age 24;3, range 20–30. Table 2 summarises relevant language background and biographical information for the advanced learner group. This does not include overall daily use, since several cumulative estimates were disproportionately high. However, all L2 learners indicated that they engaged in activities involving English for at least 90 minutes daily (e.g. writing for school, speaking to friends etc.).

Table 2: L2 learners' language background and biographical information (*n* = 18)

Learner information	Mean	Range	*SD*
Age (years)	24.3	20–30	2.7
Length of exposure to English (years)	17.4	14–26	3.1
Age of onset listening (years)	7.1	4–10	1.7
Age of onset speaking (years)	8.3	6–10	1.3
Age of onset reading (years)	8.1	5–10	1.6
Age of onset writing (years)	8.3	6–10	1.3
Daily English reading (minutes)	161.1	0–360	96.7
LexTALE score (out of 100%)	83.3	58.75–97.5	10.2

4.2 Eye-tracking

When considering eye-tracking methodology to address a specific research question, it is important to keep in mind that eye-movement data should be interpreted very carefully, since it only tells us "where on the stimulus a cognitive process operated, and possibly for how long, but not by itself *which* cognitive process is involved" (Holmqvist, Nyström, Andersson, Dewhurst, Jarodzka, & van de Weijer, 2011, p. 95–6). It should also be noted here that, whereas visual attention and eye-movements (specifically fixations, to be defined in section 4.2.1) are typically interpreted as synchronous events, visual attention may actually move slightly before the eye does (Holmqvist et al., 2011). That said, eye-tracking during reading has proven to be a valuable and reliable method for investigating online parsing procedures and for detecting readers' sensitivity to ungrammaticalities (Rayner, 1998). Rayner and Pollatsek (2006) similarly state that eye-

movements during reading "represent one of the best ways to study language comprehension processes" (p. 613). They argue that this method is relatively natural in the sense that eye-movements are not affected by an artificial, secondary task and that monitoring eye-movements during reading does not affect reading rates; readers can read at their natural pace (cf. Roberts & Siyanova-Chanturia, 2013).

Other methodologies that are commonly applied to investigate issues of language processing and comprehension – such as ERP and SPR – may be considered less natural. Both ERP and SPR typically involve a word-by-word stimulus presentation (in case of ERP, a rapid serial presentation with set presentation durations is common practice) and, in case of self-paced reading, this is rendered even more unnatural by the button presses necessary to reveal the next word or segment. Moreover, in ERP and SPR paradigms, words cannot be revisited; therefore working memory is taxed by having to retain comparatively much information. Importantly, in L2 studies, this could potentially lead to method-induced processing differences, since a word-by-word presentation may place a heavier burden on learners' working memory capacity compared to native speakers' working memory capacity. This is especially relevant when non-local dependencies are investigated, as attested by Foucart and Frenck-Mestre (2012), who found a discrepancy between ERP and eye-tracking results for violations of non-local dependencies. In eye-tracking reading studies, on the other hand, sentences can be presented in full and reading speed and the allocation of attention can be controlled by the reader: "readers can both control the amount of time they spend on a particular word and re-read a region" (Foucart & Frenck-Mestre, 2012, p. 241). Another "potential advantage that eye-movement recording holds over other online techniques is that it provides a finer-grained analysis of moment-to-moment language processing", i.e. that "reading time on words can be divided into various subcomponents" (Keating, 2009, p. 511). These components will be discussed in the next two sections.

4.2.1   Eye-movement components

The two most important eye-movement components to be discussed in this section are *fixations* and *saccades.* In fact, fixations are not movements as such; they rather reflect "the period of time when the eyes remain fairly still and new information is acquired from the visual array" (Rayner, 2009, p. 1458). During reading, the average fixation duration is 225–250 milliseconds. However,

there is much variability and short fixation durations of 50–75 ms or long fixation durations of 500–600 ms (or even longer) are not out of the ordinary (Rayner, 2009). Lexical access is also estimated to take approximately 100–300 ms to complete (e.g. Rayner & Pollatsek, 1989).

Importantly, fixation duration during reading has been shown to be affected by several factors, some of which are especially relevant to the present study. These factors include word frequency, word familiarity, age of acquisition, number of meanings (polysemy), morphology, contextual constraints, and plausibility (Clifton, Staub, & Rayner, 2007). Models of eye-movement control in reading, such as *the E-Z Reader* model developed by Reichle, Rayner, and Pollatsek (2003), typically take these properties into account (i.e. especially the ones for which there are very robust findings, such as lexical frequency). Rayner and Duffy (1986) and Inhoff and Rayner (1986) found that higher frequency words generally induce fewer fixations and shorter fixation durations compared to lower frequency ones. Function words (which are typically short and highly frequent) are often skipped during reading (e.g. Dussias, 2010; Rayner, 2009; Roberts, 2012; Schotter, Angele, & Rayner, 2012). Similar effects have been found for word familiarity, where more familiar words typically attracted shorter fixations (e.g. Juhasz & Rayner, 2003; Williams & Morris, 2004).

The polysemy of words, i.e. the characteristic that words may have several meanings, has also been shown to affect fixation duration. When words with two equally likely meanings appear in a neutral sentential context, readers generally look at them longer than they do at words that have one more dominant meaning. If, in the latter case, the sentence context reveals that a subordinate meaning was intended, fixation duration and regression probability have been shown to increase (e.g. Sereno, O' Donnell, & Rayner, 2006). Furthermore, if words are highly predictable from their previous sentential context, they are more likely to be skipped or to induce shorter fixation durations (e.g. Balota, Pollatsek, & Rayner, 1985; Rayner & Well, 1996). Moreover, Rayner, Warren, Juhasz, & Liversedge (2004) found that contextually plausible words induce comparatively shorter and less fixations. The present study aims to control for lexical frequency (and familiarity), number of meanings, contextual constraints, and plausibility (see sections 4.3.2 and 4.3.4).

As opposed to fixations, saccades are rapid eye-movements. The duration of a saccade is affected by the distance moved: "a 2-deg saccade, typical of reading, takes about 30 ms" (Rayner,

2009, p. 1459). During reading, saccades take approximately 20–40 ms to complete (Rayner & Pollatsek, 2006). However, saccades moving from the end of a line of text to the next line, called *return sweeps*, usually take longer, as do forward saccades initiated after a regression (Rayner, 2009; Rayner & Pollatsek, 2006). The average saccade length during reading in English is 7–9 characters. However, there is much variability (Rayner, 2009; Rayner & Pollatsek, 2006). Particularly for reading, it is considered safe to assume that saccades are generally too fast to allow for the simultaneous intake of new information (during saccades, vision is functionally suppressed). Cognitive processing typically continues during the saccade (Holmqvist et al., 2011; Rayner, 2009).

A third important eye-movement component is the *regression*. A regression is a saccade that moves backwards in a text. In skilled readers, regressions occur about 10–15% of the time (Rayner, 2009). However, as the general difficulty of a text increases, more saccades are made (Rayner & Pollatsek, 1989). It should be noted that "most regressions are to the immediately preceding word, though when comprehension is not going well or the text is particularly difficult, more long-range regressions occur to earlier words in the text" (Rayner, 2009, p. 1460). Although regressions can indeed result from processing difficulty, they can also stem from oculomotor error, i.e. when a saccade overshoots its intended target and needs to correct by moving back (Reichle et al., 2003).

During reading, three visual regions are of interest: the *foveal region*, the *parafoveal region*, and the *peripheral region*. The first region comprises approximately 2° of visual angle around a given fixation. The second region extends from the foveal region to about 5° on both sides of a fixation. The third region lies beyond the parafoveal region. Whereas vision is acute in the foveal region, it is much less clear in the parafoveal region and it is even worse in the peripheral region (Rayner & Pollatsek, 2006). Therefore, to be able to read (to continuously place words in or close to the foveal projection), eye-movements are necessary (Rayner, 2009). The size of the *perceptual span*, the area in which useful information intake is assumed to take place, is estimated to span 3–4 characters to the left and 15–16 characters to the right of a given fixation in English texts of average complexity (Pollatsek, Rayner, & Balota, 1986). Self-evidently, not all characters are processed equally within this span.

Despite the lower visual acuity in the parafoveal region, words can in some cases be

identified, after which they may be skipped. Especially short, contextually constrained, and/or frequent words are often skipped (e.g. Rayner, 2009; Schotter et al., 2012), which has led to the assumption that "the words that are skipped are processed on the fixation prior to the skip (and after the skip to some extent)" (Rayner, 2009, p. 1461). Additionally, Schotter et al. (2012) found evidence that readers are sensitive to the phonological, morphological, and syntactic (and in some cases semantic) properties encoded in characters and words in both the foveal and the parafoveal region. Longer (and less contextually plausible or less frequent) words are unlikely to be fully identified and processed when projected in the parafoveal region. However, they may be subject to a *parafoveal processing benefit*, i.e. their processing may be facilitated (Roberts & Siyanova-Chanturia, 2013). Another relevant phenomenon is *spillover*. The processing of a given word may spill over to the next word, i.e. fixation durations on this word may be inflated, for example as an effect of low lexical frequency (Rayner & Duffy, 1986). The eye is then programmed to move before the word in question is fully parsed.

4.2.2 Eye-movement measures in reading research

The components discussed above form the basis of the most important eye-movement measures in the field of reading research. There is some terminological inconsistency – certain measures go by different labels in different studies (see also section 4.5.2). Measures that are commonly reported include *first fixation duration*, *first-pass reading time* or *gaze duration*, *second-pass reading time*, *total reading time*, *go-past time* or *regression path duration*, *first-pass regression probability*, *rereading*, and *fixation count* (for an overview, see e.g. Roberts & Siyanova-Chanturia, 2013). These eye-movement measures are usually subdivided into *early* measures (first fixation duration and first-pass reading time) and *late* measures (the other time-based measures), although the boundary between the two is not always clearly defined (Roberts & Siyanova-Chanturia, 2013). Early measures are typically taken to be indicative of "early processes in the comprehension of a text, such as lexical access and early integration of information", whereas late measures "are believed to be sensitive to later processes associated with comprehension of a text, such as information reanalysis, discourse integration, and recovery from processing difficulties" (Roberts & Siyanova-Chanturia, 2013, p. 217).

The present study will report first pass-reading time, total reading time, and first-pass regression probability (cf. Pearlmutter et al., 1999). First-pass reading time refers to the sum of all

fixation durations in a region of interest from when it is first encountered, until it is exited either to the left or to the right, and is taken to be indicative of initial processing (i.e. inflated first-pass reading times could be indicative of early processing difficulties). Total reading time is a more inclusive measure and it is defined as the sum of all fixation durations in a region of interest, reflecting both early and late processes (i.e. initial processing difficulties and the recovery from processing difficulties). First-pass regression probability reflects the percentage of backward saccades from a specific region of interest that is fixated during first-pass reading. This measure is also thought to be indicative of increased processing load, i.e. regressions may be launched to resolve specific processing difficulties, although it is arguably somewhat coarser than reading time measures (i.e. in the sense that the difficulties likely need to be of a certain nature and magnitude to trigger this type of reanalysis).

The exact regions of interest and the adopted eye-movement measures (including two less common additional measures) and their interpretation will be discussed in more detail in sections 4.5.1 and 4.5.2, respectively.

4.3   Design

4.3.1   The reading task

The materials for the online reading task consisted of 60 experimental sentences in six conditions – thus adding up to a total of 360 experimental sentences – and 40 filler sentences. Six lists were created such that each of the 60 experimental sentences was seen in all six conditions, but only in one condition per participant. This preference for a mixed-lists design over a full within-subjects design will be motivated below. Each of the six lists contained 100 sentences in total (60 experimental sentences and 40 fillers) and ten items in a given condition. The six experimental lists were created according to a balanced Latin square design and a randomised list of sentences was generated for each participant, in which no more than two sentences belonging to the same condition were presented consecutively. The sentences were divided into blocks of 25 trials so participants could take short breaks between blocks.

Firstly, this design was employed to ensure that differences in reading patterns and processing costs for critical items would not be caused by participants' previous encounters with the constructions in question. In general, it is conceivable that participants, upon seeing sentences repeated may read later versions differently (e.g. less carefully) because of their familiarity with the

structures. Self-evidently, this is where practical solutions such as counterbalancing come into play. This technique could be adopted to counter order effects to some extent. However, the subtle physical differences between the conditions in the present study would likely amplify familiarity effects even more and therefore, counterbalancing might not be an adequate solution. More specifically, the experimental manipulations involve varying the grammatical number of the head noun, the local noun, and the past-tense copula. Pluralising the nouns involves adding *-s* (or alternatively *-es*) to their singular forms and pluralising the past-tense copula entails moving from *was* to *were*. These are subtle changes indeed (though not necessarily in a semantic sense). On a practical note, even partial counterbalancing would have been challenging considering the number of conditions in the present study. In a mixed-lists design, these problems are largely overcome.

Secondly, a mixed-lists design leaves room for a higher number of different experimental items, which is advantageous for several reasons. The effect of an inherently flawed item will even out more when it is part of a larger set (and excluding it from analyses, if necessary, will make less of a difference). Moreover, the generalised mixed effects models that were fitted to the data (see section 4.5.3) assume item sampling from a normal distribution. When the number of different items is higher, it is more likely that the data sample's distribution will approach normality and that random effects associated with by-item variation will cancel out (i.e. average near zero). Additionally, it may be presumed that it is easier for participants to stay engaged with the task when there is more variation and when the number of trials is lower. The potential for fatigue may thus be decreased. In a full within-subjects design, especially when the number of conditions is fairly high, the number of trials increases rapidly when the number of different experimental items increases. That is, were this study to be carried out with 20 experimental sentences in such a design, this would already add up to 120 experimental trials. In contrast, a mixed-lists design with three times as many different experimental items (60) consists of that exact number of experimental trials (excluding fillers). For these reasons, a mixed-lists design was chosen over a full within-subjects design. It should however be noted that mixed lists designs do have the disadvantage that they induce more item-related and inter-participant variance compared to full within-subjects designs, which may decrease the statistical power (i.e. obscure effects present in the data). Moreover, despite the disadvantages connected to higher numbers of items in full

within-subjects designs, more items do tend to increase the statistical power.

The design of the present study was factorial. Four two-level factors were varied. On top of the between-subjects factor group (with levels *native speaker* and *learner*), three within-subject factors were varied. These factors were head noun number (A: *singular* or *plural*), the relation between head noun number and local noun number (B: *match* or *mismatch*), and grammaticality (C: *grammatical* or *ungrammatical*). Since the head noun determines the grammatical number of the complex subject irrespective of the grammatical number of the local noun, (un)grammaticality is determined by (mis)matches between the grammatical number of this noun and the past-tense copula. These relations are illustrated in (8).

(8) The **editor/s** of the **newspaper/s** **was/were** admired by the staff (…)
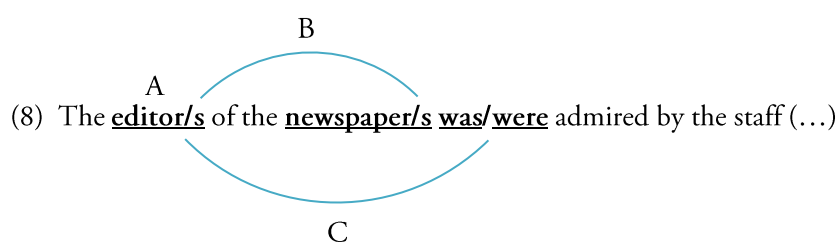
A full 2 × 2 × 2 × 2 factorial design would have resulted in sixteen different conditions and sixteen experimental lists (and it would have necessitated a higher number of participants, which did not seem realisable within the scope of this master's thesis project). However, the factors were not fully exhausted; the two plural head noun ungrammatical conditions were not included, echoing Pearlmutter et al. (1999). For these reasons, the present study will focus only on twelve out of sixteen conditions, divided over six experimental lists. At the sentence level, the six conditions that were included were as follows (9):

(9)
   a. The **editor** of the **newspaper** **was** admired by the staff… (*singular, match, grammatical*)

   b. The **editor** of the **newspapers** **was** admired by the staff… (*singular, mismatch, grammatical*)

   c. *The **editor** of the **newspaper** **were** admired by the staff… (*singular, match, ungrammatical*)

   d. *The **editor** of the **newspapers** **were** admired by the staff… (*singular, mismatch, ungrammatical*)

   e. The **editors** of the **newspapers** **were** admired by the staff… (*plural, match, grammatical*)

   f. The **editors** of the **newspaper** **were** admired by the staff… (*plural, mismatch, grammatical*)

…, especially for the excellent time management that was needed to get the job done successfully on a daily basis (sentence complement for each of the conditions).

Table 3 lists the different conditions and abbreviations for each of them. For ease of interpretation, the six conditions will henceforth sometimes be referred to as a: SSS (<u>S</u>ingular head noun, <u>S</u>ingular local noun, <u>S</u>ingular copula verb), b: SPS (<u>S</u>ingular head noun, <u>P</u>lural local noun, <u>S</u>ingular copula verb), c: *SSP, d: *SPP, e: PPP, and f: PSP.

Table 3: The different conditions and their abbreviations

Condition	Head noun	Local noun	Copula	Abbreviation
a	editor (S)	newspaper (S)	was (S)	SSS
b	editor (S)	newspapers (P)	was (S)	SPS
c	editor (S)	newspaper (S)	were (P)	*SSP
d	editor (S)	newspapers (P)	were (P)	*SPP
e	editors (P)	newspapers (P)	were (P)	PPP
f	editors (P)	newspaper (S)	were (P)	PSP

4.3.2 Materials for the reading task

One out of four structures used in Pearlmutter et al. (1999) was selected and adapted for the present study. Two of the other structures were found to be less suitable because they included either a preposition or an article in the critical word position immediately following the past-tense copula. Both prepositions and articles are typically short and highly frequent function words, which generally induce comparatively short fixation durations and which are even likely to be skipped altogether (e.g. Dussias, 2010; Rayner, 2009; Roberts, 2012; Schotter et al., 2012). Since it was expected that the copula verb itself would be skipped rather often (cf. Pearlmutter et al., 1999), a slightly longer spillover region was required. Therefore, the other two structures, with a past-participle verb or a nonparticipial adjective in critical word position, were both found to be usable. In principle, only the passive construction with the past-participle verb was adopted. This was done mainly to allow for generalisations, but also for a more practical reason. That is, only non-cognate words (i.e. words of dissimilar origin and form) were included in critical word positions, constraining stimulus generation.

To the target structure (complex subject, past-tense copula, and past-participle verb), a sentence completion, varying somewhat in length between items, was added. Importantly, this complement was longer in the present study than in Pearlmutter et al. (1999). This was done mainly to make the read somewhat more natural and to obscure the experimental manipulation to a larger extent.

Several word-level considerations were of importance for the design of the stimuli. First of all, it was necessary to control for the lexical frequency of the critical items. Although it may be expected that advanced learners have acquired numerous less frequent words, there may be a lot of individual variation. Therefore, the aim was to include only highly frequent words in critical word positions (and in neighbouring positions, taking into account possible parafoveal processing and spillover effects). To this end, a list of the top 5,000 most frequent words in American English compiled by the Corpus of Contemporary American English (COCA) was employed. This is the largest representative corpus of spoken and written English (containing 450 million words) as well as the most up-to-date one (it was last updated in 2012). From this frequency list, all nouns and verbs were extracted and ranked. Next, any homographs and all cognates were filtered out (cf. Foucart & Frenck-Mestre, 2012; Keating, 2009), since homographs (or 'false friends') may confuse learners and since L2 learners have been shown to access cognates faster than non-cognates (e.g. De Groot, Borgwaldt, Bos, & van den Eijnden, 2002; Dufour & Kroll, 1995). To this end, a native speaker of Swedish with a professional interest in etymology marked all non-cognates. In case of intended inclusion in an experimental sentence, the non-cognate status of the word was verified in online standard and etymological dictionaries.

Since the structure of the experimental sentences was fairly rigid and since inherent properties of nouns and verbs can be structurally and semantically restrictive in themselves, many of these nouns and a number of verbs were found to be unsuitable for inclusion. For example, nouns such as *news*, *love*, and *growth* are uncountable, therefore they cannot be pluralised. Moreover, nouns such as *species* have identical morphological forms in their singular and plural versions. In case of for example *behaviour* and *pressure*, both countable and uncountable versions of the nouns exist, but the countable version is more specialised or less common and therefore less natural when embedded in this structure. Furthermore, several nouns such as *size*, *shape*, and *price* are perfectly suitable structurally; however in this particular construction their plural versions are not necessarily sensible in a semantic sense. As an illustration, the combination *the sizes of the shoe* conflicts strongly with our world knowledge that tells us that a shoe does not usually have more than one size. The properties of the listed verbs were found to be much less restrictive, yet past-participle forms of intransitive verbs, such as *seemed*, *wondered*, *sounded*, and *tended*, were found to be unsuitable for this passive construction (e.g. **was/were seemed*).

To establish which nouns combined well semantically, the COCA 'list' search option was employed. This function lists the most common collocates for an entered word or phrase. Oftentimes the best fits were necessarily excluded because of their cognate status or because they were structurally or semantically unsuitable for specific conditions. In these cases, lower frequency collocates (or nouns that did not appear on the list of collocates at all) were selected. Care was taken to combine noun phrases in semantically natural ways, since contextually unexpected or implausible words have been shown to increase fixation duration of the critical word or of the word following the critical word, depending on the degree of unexpectedness or implausibility (e.g. Ashby, Rayner, & Clifton, 2005; Rayner et al., 2004). The semantic suitability of the selected past-participle verbs was also verified through this search function.

Finally, 120 unique nouns and 54 unique verbs were included. Six verbs (*discovered, encouraged, expected, improved, intended,* and *left*) were reused once. The learners' knowledge of nouns and verbs that ranked below the cut-off point of 2,000 was tested in a multiple choice task that will be described in section 4.3.4. Self-evidently, lexical frequency per se is too generalised in the sense that it may obscure the fact that a word may be polysemous, i.e. have several meanings, including less frequent and more specialised ones. How this problem was dealt with will be addressed in section 4.3.4.

Critical nouns that involved character changes and/or addition of *-es* instead of *-s* when pluralised (e.g. *enemy - enemies*) were avoided as much as possible. Furthermore, words and phrases that are especially frequent in American English (e.g. *Republican* and *movie*) but not in British English (or vice versa) were not included. In case of doubt, the entry for the word (or phrase) in question was looked up in the Cambridge Dictionaries Online series (http://dictionary.cambridge.org/), which provides information about American and British uses. Moreover, the stimuli were spelled and constructed according to the rules of British English (e.g. *rumour, organised,* and *travellers*), since the native speaker group consisted of native speakers of British English.

Finally, a rather varied set of 40 grammatically correct filler sentences was created. Some filler items were structurally similar to the experimental sentences, but most of them were strongly different in terms of construction and even tense. This was done to obscure the experimental manipulations to some extent as well as to prevent experimental fatigue.

Importantly, all experimental items and filler sentences were accompanied by simple (but sometimes rather unexpected) Yes/No content questions to ensure that the participants would stay on task. The number of questions requiring a positive answer and a negative answer was balanced. A sentence comprehension task was deemed to be more appropriate than a grammaticality or acceptability judgement task, since the latter might directly alert the participants to the experimental manipulations. Importantly, the comprehension questions never targeted the experimental manipulations. The accuracy results of the comprehension task were used to ensure that participants had attended to the stimuli, but they did not affect the outcome of the study. The complete list of experimental and filler sentences and accompanying questions can be found in Appendix I. On completion, all experimental and filler sentences and content questions were checked by a native speaker of British English with a professional background in editing, in order to verify their accuracy and appropriateness.

4.3.3 LexTALE: lexical test for advanced learners of English

The LexTALE vocabulary test (Lemhöfer & Broersma, 2012) was administered to the learners. The LexTALE is an unspeeded lexical decision task in which participants simply have to decide whether the strings of letters they are presented with are real English words. The test consists of a set of 40 real words that vary strongly in lexical frequency and 20 non-words that structurally resemble real English words. It has been shown to be a good indicator of English vocabulary knowledge and a fair indicator of general English proficiency. Furthermore, it was established that it is a better and a more consistent measure of proficiency compared to self-ratings (Lemhöfer & Broersma, 2012), and that it takes much less time to complete compared to extensive proficiency tests such as the Quick Placement Test (QPT) and the Test of English for International Communication (TOEIC) test. Therefore, the LexTALE was considered the most suitable proficiency task for this study. The task was implemented and scored according to the guidelines stated on the LexTALE webpage (http://www.lextale.com/). The percentages correct in each of the categories (i.e. real words that were recognised and non-words that were rejected) were calculated, added together, and divided by two (see section 5.1.2, example (11)).

4.3.4   Multiple choice test for the advanced learner group

A multiple choice task was designed to test the learners' knowledge of less frequent critical words (rank 2,000 in the COCA top 5,000 frequency list was defined as the cut-off point). However, as discussed, lexical frequency per se is not a straightforward indicator of what advanced learners (or native speakers) may be expected to know. Importantly, words may be polysemous and specific senses may be much less frequent and/or more specialised. The Cambridge Dictionaries Online series was employed to control for this to a certain extent. In many of their entries, the Cambridge dictionaries offer information about the CEFR levels of words and phrases (A1–C2), as categorised by the *English Vocabulary Profile*. This classification is focused on what learners 'know' rather than what they 'should know' (Capel, 2010) and it is "underpinned by up-to-date corpus evidence, including the 50-million word Cambridge Learner Corpus and the 1.2-billion word Cambridge English Corpus of first language use" (Capel, 2012, p. 1).

The English Vocabulary Profile classifies levels A1–2 and B1–2 as 'basic user' and 'independent user' levels, respectively, whereas the C1–2 levels are defined as 'proficient user' levels. Although it may be expected that advanced learners know many words up to and including the C1 level, critical words with a C1 (and in one case a C2) label were incorporated in the multiple choice test. The labels for the critical words had to be decided upon first. In case of polysemy, this was done by matching the word as closely as possible to one of its dictionary definitions, taking into account its context in the experimental sentence. In most cases, a critical item was included either because of its frequency status or because of its CEFR-level, but sometimes lower frequency and C1 status overlapped. In the end, 31 items (23 nouns and 8 verbs) were included in the test.

In the multiple choice test, a definition of a word – which was adapted from definitions in the Cambridge Learner's and British English dictionaries – was presented, and participants were asked to choose the correct word from four alternatives. One of these alternatives was the target word, one option was semantically strongly related (e.g. near-synonymous), another was somewhat semantically related, and one was wholly unrelated. The highly and somewhat related options were found by entering the search string [= *x*] in the COCA, which returns synonyms and near-synonyms, and through the Cambridge Dictionaries' Thesaurus function. In some cases, a native speaker of British English was consulted. The possible answers were organised

according to a Latin square design, after which the trial order was randomised in the experimental software (see section 4.4). Since one of the experimental sentences was replaced last-minute, the distribution of the answers was slightly skewed (i.e. the sentence in question contained three critical words that were originally included in the test).

It could be argued that the correct answers were primed or that specific words were learned because they appeared in the experimental sentences, in which case correct answers could be false reflections of the participants' previous knowledge. However, since the online reading experiment consisted of 100, fairly long, sentences (and accompanying questions) and since the LexTALE was administered in between, it is likely that not many critical items (and any newly assigned meanings) were retained in the participants' short-term memory. Moreover, when possible, nouns and verbs that appeared elsewhere in experimental or filler sentences were included in the test as alternative answers.

4.3.5 Language history questionnaire: LHQ 2.0

An extensive language history questionnaire, LHQ 2.0, designed by Li, Zang, Sai, and Puls (2014), was used for the purpose of this study. The authors based the generic questionnaire on the most commonly asked questions in previously published studies with second language learners or bilinguals. The questionnaire includes blocks of questions on general participant background (e.g. age and profession), language history, language usage, and language dominance. The web-based LHQ 2.0 interface is convenient, and the privacy of the participants is well-protected (the data can only be accessed by the researcher through a password-secured system).

4.4 Apparatus and procedure

During the reading experiment, eye movements were monitored with the EyeLink 1000 tower-mounted system designed by SR Research. Viewing was binocular and eye-movements were recorded from the right eye, at a sampling rate of 1000 Hz. The tracker was interfaced with a PC controlling the display of the stimuli and data storage. Sentences and comprehension questions were displayed individually on a 24 inch high resolution display monitor (ASUS VG248QE). The selected resolution was 1920 × 1080 and the refresh rate was 144 Hz. Both sentences and questions were double-spaced and appeared in multiple (2–3) lines in black against a white background. Sentences were presented in Courier New font, which was found to be suitable

because of the equal width of all of its characters (therefore, there was a one-to-one relationship between the number of characters per word and word length). Each individual character spanned approximately 0.4° × 0.5° of the visual angle; at a viewing distance of 625 mm. Anti-aliasing settings were applied to all text resources to ensure that the read was as smooth and pleasant as possible. Important critical words were never presented at the beginning or at the end of a line, to avoid corrective fixations and sentence wrap-up effects (Just & Carpenter, 1980; Rayner, Kambe, & Duffy, 2000). The experiment was designed in Experiment Builder (version 1.10.1025) and the data was extracted with DataViewer (version 1.11.900), created by SR Research. The relevant dependent measures were extracted through the 'Get Reading Measures' application, which was also developed by SR Research, especially for the purpose of reading research.

Before the experiment started, participants were asked to read the first part of a consent form, which provided a broad outline of the experimental procedure and the general aim of the study. It stated the participants' right to withdraw at any time, as well as their right to have their data destroyed in case of dissatisfaction with the purpose of the study. Only after the full experiment were they asked to read the second part of the consent form containing important post-experiment information, and to sign it. After having read the pre-experiment part of the consent form, participants were asked to take place behind the eye-tracker, to rest their foreheads against the forehead rest and to place their chin on the chin rest. The table height and the height and protrusion of the chinrest were adjusted to fit their individual needs. Moreover, the camera settings were adjusted to arrive at an optimal image of their pupil and the corneal reflection. After this, a 9-point calibration was performed, followed by a validation (this procedure was repeated if necessary). Participants read instructions as well as three practice trials, during which they were assisted by the experimenter.

Each experimental sentence was preceded by a fixation cross that was presented for 1000 milliseconds before stimulus onset. The fixation target was positioned midway between the far left end of the display screen and the sentence initial character. Participants were instructed to focus on it before a new sentence appeared. They were asked to read the sentences for meaning, to be able to answer simple comprehension questions directly following each sentence. Participants pressed space when they were done reading a sentence, after which a question appeared from the same starting position. Questions were answered using the Yes/No buttons on

the keyboard (the left and right arrow, respectively), after which a new fixation cross appeared. Participants were aware that sentences and questions timed out after 30 seconds, but they were told not to feel rushed and to read at a normal pace. The sentences and questions were divided in four blocks of 25, with 30 seconds pauses between blocks (or shorter pauses, if participants wished to continue before the pause had timed out).

After the online experiment, the native British participants continued with the language history questionnaire, which was filled out on the experiment leader's private laptop. The Swedish participants continued with the LexTALE and the multiple choice task, respectively. This sequence was chosen to alternate between more and less demanding tasks (to prevent experimental fatigue) as well as to increase the validity of the multiple choice test. This order of tasks was piloted multiple times before the actual recording phase. Participants were told they could move freely again after the reading experiment. The LexTALE items were presented in the very centre of the monitor and the same Yes/No buttons were used for this task. In the ensuing multiple choice task, participants responded to the multiple choice items with the a, b, c, and d buttons on the keyboard. After these additional tasks, the Swedish participants were also asked to fill out the language history questionnaire on the researcher's private laptop. The complete session lasted approximately 35–40 minutes for the British participants, and 65–70 minutes for the Swedish participants.

4.5   Methods of analysis

4.5.1   Regions of interest

Echoing Pearlmutter et al. (1999), the *verb region* was defined as the main region of interest. This region consisted of the past-tense copula and the past-participle verb combined. In their analyses, Pearlmutter et al. (1999) decided to focus mainly on the verb region rather than on the past-tense copula and the following word individually, because the effects of the manipulations were found to be more robust for the region as a whole. Firstly, and not surprisingly, it was found that the copula was often skipped during reading (i.e. the copula verb is a short, highly frequent function word). For this reason, and because of the general possibility of spillover effects (Pearlmutter et al, 1999; Rayner & Duffy, 1986; Wagers et al., 2009), the verb region was also defined as the main region of interest in the present study. Apart from that, and in line with Pearlmutter et al. (1999), the two individual verbs and the two nouns that were part of the complex subject were defined as

regions of interest. The two nouns and the copula verb were the regions manipulated for grammatical number. The exact regions of interest are presented in (10); 1: head noun, 2: local noun, 3: copula verb, 4: past-participle verb, 5: verb region.

(10)　　　The | editor | of the | newspaper |　| was | admired | by the staff (...)
　　　　　　　　1　　　　　　2　　　　　3　　　4
　　　　　　　　　　　　　　　　　　　　　　　5
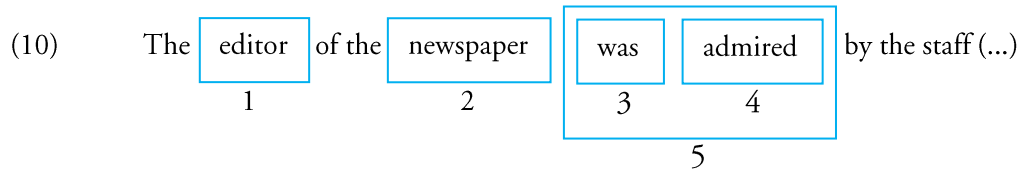
4.5.2　Dependent measures

Before specifying the dependent measures that were investigated in the present study, it should be noted once more that reading research, even though it is a very well-researched, well-constrained, and sophisticated research domain (Holmqvist et al., 2011), is characterised by a terminology that can be rather inconsistent at times. This clearly applies more strongly to some dependent measures than to others, however. To avoid confusion about which dependent measures this study investigated exactly, clear labels and definitions will be provided below.

4.5.2.1　First-pass reading time/gaze duration

The first and most important dependent measure was *first-pass reading time* or *gaze duration*. This measure, which indexes early processing difficulties, was expected to most clearly reflect qualitative between-group processing similarities or differences. In the SR Research application used to extract the dependent measures, this measure is referred to as *first-pass duration*. This measure was defined as the sum of all fixation durations in a region of interest from when it was first entered until it was exited either through a regressive or a progressive saccade, typically ending in a new fixation (e.g. Roberts & Siyanova-Chanturia, 2013). Importantly, this measure received an N/A (not available) value if either of the two critical nouns or the region itself was skipped during first-pass reading (if a word or a text region to the right of the region of interest was fixated before the region of interest was first fixated). This was done to ensure that both critical nouns were read. Percentages of excluded data will be reported when relevant in the results chapter (see section 5.2). First-pass reading time has been found to be sensitive to semantic and syntactic anomalies (Rayner et al., 2004). In the present study, mismatches and anomalies were expected to inflate first-pass durations in the verb region for the native speakers.

4.5.2.2 Total reading time

The second dependent measure was *total reading time* (*total duration* in the SR Research application). This measure refers to the sum of all fixation durations in a specified region of interest. As it includes first-pass reading time, as well as later reading times, it is thought to reflect both initial processing and time spent recovering from processing difficulties (Roberts & Siyanova-Chanturia, 2013). Liversedge, Paterson, and Pickering (1998) proposed that, if no effects are observed for earlier measures such as first fixation duration and first-pass reading time, but effects are found for total duration, this may be indicative of late processing effects for the manipulation. Total reading times were investigated for the verb region as well as for all individual regions of interest. If the region of interest was skipped altogether during a specific trial, this measure received an N/A value. If either of the critical nouns was skipped before the verb region (or either of the individual verbs) was read for the first time, this measure also received an N/A value. In the present study, the total reading times for the verb region (and the individual verbs) were expected to be inflated for the native speakers in ungrammatical conditions compared to grammatical ones, reflecting the resolution of the processing difficulty arising in the verb region. Importantly, learners were expected to reread sentences (or parts of sentences) more often in general, irrespective of for example grammaticality. Therefore, no very clear effects were expected for this measure for the learner group.

4.5.2.3 First-pass regression probability

The third dependent measure was *first-pass regression probability*. This measure reflects the percentage of backward regressions from a specific region of interest that is fixated during first-pass reading. It has been argued to reflect increased processing load (e.g. Vasishth & Drenhaus, 2011). In the present study, the measure was calculated for the verb region. It was expected that, if a seeming or actual anomaly would be encountered, this would increase the first-pass regression probability for the native speakers. If first-pass reading of the verb region resulted in a backward regression, it received a score of one. If it resulted in a progressive saccade, it was scored zero, and if the region was skipped during first-pass reading (or if either of the two critical nouns was not read during first-pass reading), it was reported as a missing (N/A) value. From these values, the regression probabilities for each condition were calculated. It should be noted that Pearlmutter et al. (1999) investigated first-pass regression probability as well as total regression probability,

whereas the present study only investigated first-pass regression probability. This was done to make the comparison between the two groups as valid as possible (i.e. the learners were generally expected to reread sentences more often).

4.5.2.4 Final landing sites for first-pass regressions

The final landing sites for first-pass regressions initiated from the verb region were considered to be of interest (cf. Keating, 2009; Pearlmutter et al., 1999), because they could potentially reflect where the processing difficulty was resolved by the reader. That is, for the native speakers, the complex subject's head noun and to a lesser extent the local noun were expected to be relevant landing sites for first-pass regressions from the verb region, since these were the regions where the seeming or actual anomalies could potentially be resolved. However, it should be noted that Pearlmutter et al. (1999) found that "targeting of regressions was not strongly controlled by syntactic processes" (p. 441). Instead, the 1999 study found an effect of linear distance, i.e. the local noun was a much more frequent landing site than the other four words preceding the verb region. The head noun was regressed to only slightly more often than would be expected in such a linear distance account, while it should in fact be the most relevant landing site for the resolution of processing difficulties. It should be mentioned that the said study investigated regressions from all passes through the verb region, whereas the present study focused only on first-pass regressions.

4.5.2.5 Probability of skipping the copula during first-pass reading

The choice of the verb region as the main region of interest was motivated primarily by the copula skipping phenomenon common for native speakers. The present study also investigated the probability of skipping the copula during first-pass reading for both groups. If the copula verb was skipped during first-pass reading, it received a value of 1; if it was fixated, it received a zero value. Defining 'skipping' rather than 'fixating' as a 'success' may seem rather counterintuitive, but it results in a very clear visual (see section 5.2.2, Figure 6). This measure was expected to reveal interesting, general processing differences or similarities between the two groups. We hypothesised that the advanced L2 learners would fixate the copula more often compared to the native speakers. It is conceivable that L2 learners generally need to read somewhat more carefully to understand the gist of a sentence, whereas for native speakers it may often suffice to just 'scan'

a sentence (this depends on the sentence's level of complexity). The setting of a sentence comprehension experiment likely motivates both native speakers and L2 learners to read more carefully; however, this may apply more strongly to L2 learners (causing them to fixate comparatively more words). Furthermore, native speakers may be subject to more specific processing benefits allowing them for example to process information more incrementally, to predict upcoming information more successfully, and to process more information parafoveally (which could induce higher skipping rates).

4.5.3   Statistical analyses

First, a fixation report was created in DataViewer, after which the 'Get Reading Measures' application extracted the selected measures. This was done for all specified regions of interest, consisting either of individual words (the head noun, the local noun, the copula, and the past-participle verb) or two words combined in the verb region (the copula and the past-participle verb). The raw reading times were then trimmed and transformed, and analysed with linear mixed effects models with random intercepts and slopes.

Before linear mixed-effects regression analyses were performed, single fixations shorter than 50 ms were removed from the data. After that, reading times shorter or longer than 2.5 standard deviations from the mean of each participant for each condition were excluded. Two operations were performed to normalise the data. Firstly, the raw reading times were log-transformed (their natural logarithms were calculated). Secondly, the log-transformed reading times were transformed into residual reading times (Ferreira & Clifton, 1986; for discussion, see Trueswell, Tanenhaus, & Garnsey, 1994). To this end, a linear regression equation estimating reading times from word or region length was fitted for each participant (using all experimental sentences). For each region of interest, the reading times predicted by the individual regression equations were subtracted from the actual reading times to obtain the residual reading times. Residual reading times are more normally distributed compared to raw reading times and they adjust for differences in word or region length across conditions, as well as for overall differences in participants' reading rates. Analyses of log-transformed raw reading times will be presented alongside those of residual reading times. Binary variables (i.e. first-pass regression from the verb region and copula skipping) were logit-transformed to arrive at probabilities.

A linear mixed-effects model was fitted to the data. This model included fixed-effect terms

for the between-subjects factor group (native speaker, learner) and the within-subjects factor condition (a: SSS, b: SPS, c: *SSP, d: *SPP, e: PPP, f: PSP; for the meanings of these abbreviations, see section 4.3.1, Table 3) and crossed random effects for participant and item. The model contained both random intercepts (i.e. mean values for each participant and each item were allowed to vary) and random slopes (i.e. slopes for each participant and each item were allowed to vary, since it may be expected that different participants and different items have different degrees of sensitivity to the manipulation at hand) and a correlation between these two (cf. Barr, Levy, Scheepers, & Tily, 2013). Since participants were repeatedly measured on sentences in each condition, participant random slopes for condition were included. Moreover, since each sentence appeared in different conditions, item random slopes for condition were included. An interaction term was included between the model's two main fixed effects group and condition. Other potentially relevant independent variables such as centered age and the centered length of the word preceding the region of interest (i.e. in case of the verb region) were investigated. Only if the addition of a predictor variable significantly improved model fit was it retained as a fixed effect.

Linear mixed-effects models have several advantages over more traditional techniques used in L2 research (i.e. ANOVAs and t-tests). Firstly, they allow a simultaneous account for random effects for both participants and items (Cunnings, 2012). Moreover, several independent variables of interest – both of categorical and of continuous nature – can be included as fixed effects in such models. Likewise, both categorical (i.e. binary variables, such as first-pass regression from the verb region in the present study) and continuous dependent variables (e.g. total reading time in ms) can be investigated. Finally, mixed-effects models are robust against missing data (and otherwise unbalanced datasets), which are a common problem in L2 research.

All statistical analyses were performed in R version 3.1.1 (R Core Team, 2013) and the mixed-effects regressions were carried out using the lme4 package, version 1.1-7 (Bates, Maechler, Bolker, & Walker, 2014).

5 Results

5.1 Offline and background tasks

5.1.1 Reading comprehension accuracy

Responses to the comprehension questions following each of the experimental sentences in the reading experiment were recorded and mean accuracy scores were calculated for each participant and for each group. The mean accuracy score for the British native speakers was 94.8%, *SD* = 3.49 (range: 88–100%) and for the advanced learner group 96.1%, *SD* = 3.03 (range: 90–99%). These high mean (and minimum) scores for both groups indicated that all participants had attended to and understood most of the experimental sentences and comprehension questions. As mentioned in section 4.1.1, one native British participant was excluded due to low comprehension accuracy (57%). Levene's test was applied to verify the homoscedasticity of the variances (F = 0.023, p = 0.88). An independent samples t-test indicated that the group means did not differ significantly (t(31) = 1.16, p = 0.26).

5.1.2 LexTALE accuracy

Responses to the LexTALE were scored and accuracy on this test was calculated for each participant in the learner group according to the formula presented in (11).

(11) $$\frac{((\text{number of words correct}/40) \times 100) + ((\text{number of non-words correct}/20) \times 100)}{2}$$

The mean LexTALE accuracy score was 83.3%, *SD* = 10.2% (range: 58.75–97.5%). Different graphical representations of the data revealed either one or two outliers. These included the minimum (58.75%) and in some cases the second lowest score (63.75%). Since these two participants demonstrated high reading comprehension accuracy (respectively 95% and 99%) and since they had fairly high scores on the multiple choice test (both 86.7%), they were not immediately excluded from all analyses. The other 16 LexTALE scores were roughly normally distributed.

5.1.3 Multiple choice test accuracy

Responses to the multiple choice vocabulary test were scored and an accuracy percentage was calculated for each participant. Since the multiple choice test was specifically designed for the

purpose of this study, it was not standardised. Therefore, the distribution of correct and incorrect responses for each of the items in the test was investigated before the accuracy scores were calculated. Out of 31 multiple choice test items, 13 were answered correctly by all participants and 18 definitions were matched to a wrong word (or phrase) by at least one and at most 10 participants. The distribution of the accuracy scores for these 18 items showed one clear outlier. The item in question was answered correctly only by eight out of 18 participants, likely because one of the alternative answers was too strongly synonymous with the correct answer (see Appendix I), and it was removed before the individual accuracy scores were calculated (the experimental sentence containing this word was retained for eye-tracking analyses since the word had a B2-classification, and because the flaw in the design of the multiple choice test was rather clear). The mean accuracy score for the remaining 30 items was high, 92.1%, *SD* = 3.73 (range: 86.7–100%). These results indicated that the participants generally had good knowledge of the meanings of the comparatively less frequent and/or high CEFR-level critical words used in the present study.

5.2   Eye-tracking during reading

5.2.1   Verb region

Linear mixed-effects regressions were performed on the first-pass and total reading times in this main region of interest and a logistic mixed-effects regression was applied to the first-pass regression probabilities. The native speaker group and condition a: SSS (i.e. the condition not involving any plural number marking) were chosen as the baseline, but releveling was done when necessary (i.e. selecting a different level of either of the factors or both of the factors as the baseline).

First-pass reading time. The model-estimated mean first-pass reading times (ms) for both groups are presented in Figure 1 and in Table 4. To prevent distortion (i.e. by calculating the means of the raw reading times for each condition), mean first-pass reading times were calculated by exponentiating the model's fixed effects intercept (i.e. to arrive at the mean first-pass reading times for condition a) and the sum of the intercept and each of the fixed-effects coefficients for condition (i.e. to arrive at the mean values for each of the other conditions). It should be noted that this linear mixed-effects model was applied to the trimmed and log-transformed, but non-residualised reading times (these will be presented later for comparative purposes).
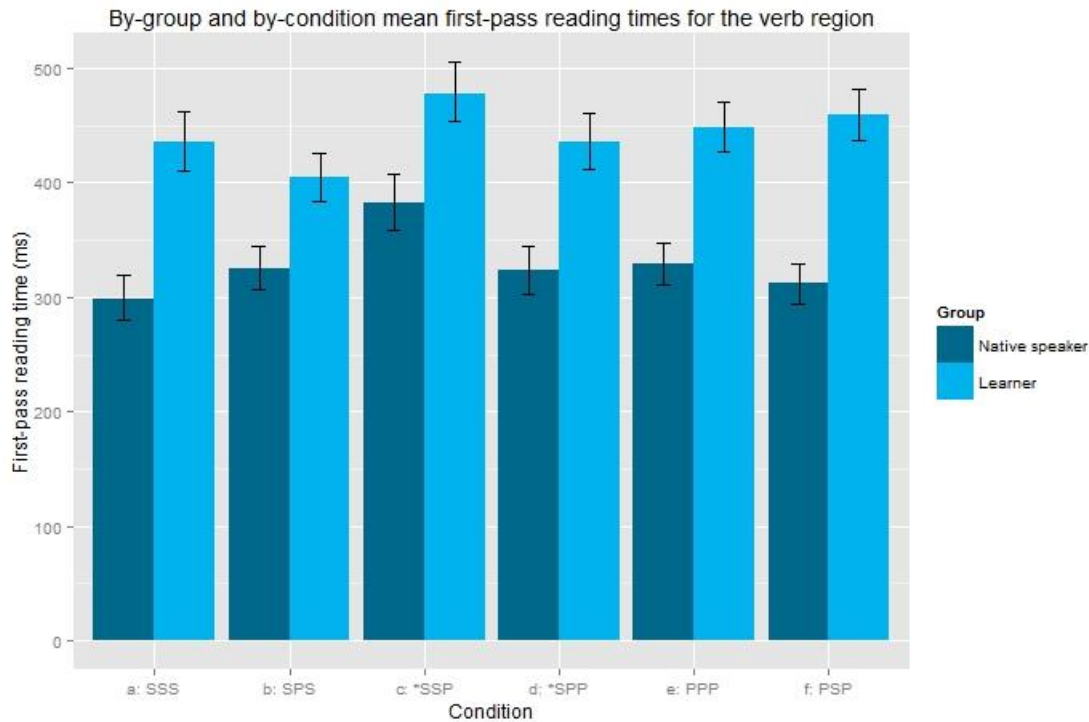
45

The error bars in Figure 1 and subsequent figures represent standard errors (SEs) for the by-group, by-condition means (for this sample of items and participants). SEs are indicative of the uncertainty around each estimate, while the SDs presented in Table 4 and subsequent tables are indicative of the variability in the data sample. Figure 1 and Table 4 give a good indication of the general patterns in the first-pass reading data (in ms), for both groups. They indicate that reading times for the learner group were generally longer, i.e. L2 learners were slower than native speakers.

Table 4: Mean first-pass reading times (ms) in the verb region

Condition	Mean first-pass reading times (ms) native speakers (difference from baseline)	SD	Mean first-pass reading times (ms) learners (difference from baseline)	SD
a: (SSS)	299	80	435	113
b: (SPS)	325 (+26)	75	405 (-30)	90
c: (*SSP)	383 (+84)	93	478 (+43)	115
d: (*SPP)	323 (+24)	80	436 (+1)	105
e: (PPP)	329 (+30)	72	448 (+13)	92
f: (PSP)	312 (+13)	67	459 (+24)	98

The regression output of the mixed-effects regression performed on the log-transformed first-pass reading times is presented in Table 5. Model outputs for the native speakers and the L2 learners are merged into the same table. The second p-value (p-res) represents a model applied to the residual reading times (the full model output can be found in Appendix II). It was included to show the effect of the length-correction/residualisation on the results.

Table 5: Regression output for the log-transformed (non-residualised) first-pass reading times

Fixed effects	Estimate (log)	Std. Error	t-value	p-value	p-res
Intercept (condition a: native speakers)	5.701	0.066	86.37	0.000 *	0.020 *
Condition b	0.083	0.058	1.43	0.164	0.006 *
Condition c	0.248	0.062	3.97	0.000 *	0.000 *
Condition d	0.076	0.063	1.21	0.237	0.170
Condition e	0.094	0.056	1.67	0.104	0.190
Condition f	0.043	0.056	0.78	0.443	0.919
Group learner	0.375	0.090	4.17	0.000 *	0.149
Age (centered)	-0.038	0.014	-2.79	0.009 *	
Length preceding noun (centered)	0.019	0.006	3.06	0.005 *	
Condition b: group learner	-0.156	0.077	-2.01	0.053	0.016 *
Condition c: group learner	-0.155	0.083	-1.87	0.071	0.016 *
Condition d: group learner	-0.075	0.081	-0.92	0.362	0.538
Condition e: group learner	-0.066	0.073	-0.91	0.371	0.449
Condition f: group learner	0.009	0.073	0.13	0.900	0.448
Intercept (condition a: learners)	6.076	0.059	103.22	0.000	0.539
Condition b	-0.073	0.052	-1.40	0.171	0.617
Condition c	0.094	0.055	1.71	0.097	0.397
Condition d	0.001	0.056	0.02	0.982	0.498
Condition e	0.028	0.048	0.57	0.570	0.703
Condition f	0.053	0.049	1.08	0.289	0.215

Estimates and standard errors are in log-transformed milliseconds. There were 1813 observations (out of a total of 1980). Data removal (i.e. when either of the two critical nouns or the verb region itself was skipped during first-pass reading) resulted in the exclusion of 159 trails (8.0%). Further trimming at 2.5 SD of the mean of each participant for each condition resulted in the removal of 8 more observations (0.4%). *P*-values were obtained through an F-test based on the Kenward-Roger approximation of degrees of freedom. The value for *p*-res was based on 1521 observations.

Table 5 confirms that first-pass reading times were indeed longer for the learners (p < .001) compared to the native speakers. Model comparisons indicated that the addition of centered age and the centered length of the word preceding the verb region significantly improved model fit. First-pass reading times in the verb region were found to decrease (p < .01) with age (i.e. older participants generally spent less time reading this region). Moreover, if the word preceding the verb region (i.e. the local noun, which also varied in length with condition) was longer (i.e. in number of characters), reading times in the verb region were also longer (p < .01). This predictor

was added as a fixed effect mainly to control for possible spillover effects (i.e. in case of a plural local noun). For the British native speaker group, a main effect was found for condition c ($p <$ .001). Releveling to this condition showed that first-pass reading times were longer in the c-condition compared to all other conditions. No main effects were found for the Swedish learners, nor were any interactions found between group and condition.

In line with Pearlmutter et al. (1999), regression-ended first passes in the verb region were then removed from the data sample; leaving 1521 observations for analyses. This was done because both fixating a specific region of interest longer as well as regressing from it upon first-pass reading may be considered to be different 'strategies' for participants to tackle processing difficulties arising in that region. Pearlmutter et al. (1999) found that first passes in the verb region that ended in a regression were substantially shorter than non-regression-ended passes; and results were statistically stronger when regression-ended first passes were excluded. Therefore, this procedure was also applied in the present study. For the native speaker group, on top of a main effect for condition c ($p <$ .001), another main effect was revealed for condition b ($p <$ .05). The c-condition was found to be more difficult than all of the other conditions, except the b-condition. For the learners, main effects were found for conditions c ($p <$ .05) and f ($p <$ .05). Moreover, an interaction between condition b and group was found ($p <$ .05), such that the learners spent comparatively little time reading the verb region in the b-condition.

When further transforming the first-pass reading times for the 1521 non-regression ended-passes into residual reading times, the pattern changed somewhat (i.e. some of the variance in the model could be attributed to region length effects). Main effects were found for the native speakers for conditions b ($p <$ .01) and c ($p <$ .001). The c-condition was again found to be more difficult than any of the other conditions, except the b-condition. Moreover, two interactions were found for the b ($p <$ .05) and c conditions ($p <$ .05), indicating that the learners spent comparatively little time reading this region in these conditions. No main effects were found for the learner group after releveling (i.e. the main effects for the c and f conditions disappeared with the residualisation and were therefore likely an artifact of region length). The outputs of the second and third linear mixed-effect regressions described above can be found in Appendix II, in Table 12 and Table 13. The p-values for the third model are also printed in Table 5.

Total reading time. The mean total reading times (ms) for each group and each condition are presented in Figure 2 and Table 6.
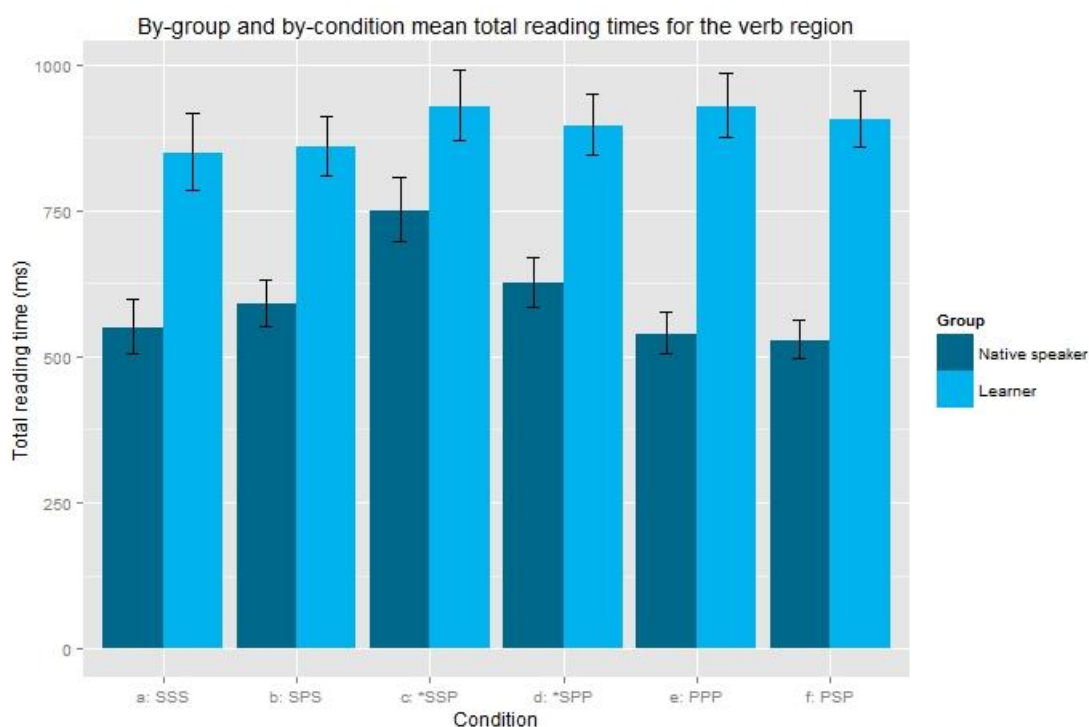
Figure 2 clearly indicates that the total reading times for the learners were generally longer than those for the native speakers. This was likely mainly due to more rereading (i.e. an independent samples t-test indicated that the group means for rereading duration differed significantly ($t(31)$ = 2.93, p < .01), but also to generally slower reading in the L2, as reflected in the first-pass reading times.

Table 6: Mean total reading times (ms) in the verb region

Condition	Mean total reading times (ms) native speakers (difference from baseline)	*SD*	Mean total reading times (ms) learners (difference from baseline)	*SD*
a: (SSS)	549	189	847	288
b: (SPS)	591 (+42)	153	858 (+11)	219
c: (*SSP)	750 (+201)	220	929 (+82)	260
d: (*SPP)	627 (+78)	162	895 (+48)	227
e: (PPP)	539 (-10)	143	929 (+82)	239
f: (PSP)	528 (-21)	131	906 (+59)	209

The regression output of the mixed-effects regression performed on the log-transformed, non-residualised total reading times is presented in Table 7. Model comparisons indicated that the addition of the centered length of the word preceding the verb region significantly improved model fit.

Fixed effects	Estimate (log)	Std. Error	t-value	p-value	p-res
Intercept (condition a: native speaker)	6.309	0.086	73.71	0.000 *	0.209
Condition b	0.073	0.066	1.11	0.275	0.187
Condition c	0.311	0.073	4.26	0.000 *	0.001 *
Condition d	0.132	0.066	2.01	0.052	0.094
Condition e	-0.020	0.067	-0.30	0.767	0.509
Condition f	-0.040	0.062	-0.65	0.518	0.126
Group learner	0.433	0.111	3.92	0.000 *	0.696
Length preceding noun (centered)	0.035	0.014	2.50	0.017 *	
Condition b: group learner	-0.060	0.079	-0.76	0.451	0.586
Condition c: group learner	-0.219	0.093	-2.34	0.025 *	0.014 *
Condition d: group learner	-0.078	0.084	-0.92	0.362	0.349
Condition e: group learner	0.112	0.084	1.34	0.188	0.172
Condition f: group learner	0.107	0.081	1.33	0.191	0.166
Intercept (condition a: learners)	6.742	0.078	86.66	0.000 *	0.389
Condition b	0.013	0.059	0.23	0.823	0.444
Condition c	0.092	0.064	1.43	0.161	0.748
Condition d	0.055	0.058	0.94	0.352	0.574
Condition e	0.092	0.059	1.57	0.126	0.267
Condition f	0.067	0.054	1.25	0.220	0.825

Estimates and standard errors are in log-transformed milliseconds. There were 1812 observations (out of a total of 1980). Apart from the 159 observations (8.0%) that were removed because of first-pass skipping, further trimming at 2.5 SD of the mean of each participant for each condition resulted in the exclusion of 9 more observations (0.5%).

Again, a main effect was found for the native speakers for condition c ($p < .001$), whereas the main effect that was found for condition b in the first-pass reading times disappeared. For the d condition, the effect approached significance ($p = 0.052$). Releveling to the c-condition showed that this condition was reliably more difficult than any of the other conditions. No main effects for condition were found for the learners. A between-group interaction was found for the c-condition ($p < .05$), indicating that the total reading time difference between the a and c conditions was significantly larger for the native speakers compared to that for the advanced learners.

When the log-transformed total reading times were further transformed into residual

reading times, the findings were highly similar. A main effect was found for the native speakers for condition c (p < .01), as well as a c-condition and group interaction (p < .05) in the same direction as that for the log-transformed, non-residualised reading times. Again, releveling to the c-condition indicated that reading times for this condition were longer than those for any of the other conditions. No main effects were found for the advanced learners. The full output of the second model can be found in Appendix II, Table 14. The p-values of that model output can also be found in Table 7.

Influence of L2 learner proficiency. To explore the effects of learner proficiency on reading times, it was investigated whether the LexTALE vocabulary and proficiency test score was a good predictor of first-pass and total reading times in the verb region. To this end, the (centered) LexTALE score was added as a fixed effect in the model applied only to the subset of advanced learners. However, it was found that the score was not a reliable predictor for either of the measures. Excluding the two Swedish participants with the lowest LexTALE scores (statistical outliers) from linear mixed-effects analyses did not affect the results in any meaningful way either (i.e. these analyses were applied to the otherwise full set of participants, including the native speakers). Finally, the individual first-pass and total reading time patterns for the five best performers on the LexTALE task (score > 90) were investigated in more detail (both graphically and through simplified linear-mixed effects analyses), but no further effects were revealed. Therefore, all further analyses were performed on the full set of participants.

First-pass regression probability. First-pass regression probability was investigated using a logit mixed-effects model with the same fixed and random effects specification that was used for the time-based measures discussed above. Main effects were found for the native speaker group for conditions e (p < .05) and f (p < .01), such that the probability of a first-pass regression from the verb region was smaller for these conditions compared to that for the baseline condition. For the learners, no significant main effects were found. Interactions were found for conditions e (p < .01) and f (p < .05), indicating that the native speakers regressed back from the verb region comparatively little in these conditions. The output of the logit mixed-effects regression can be found in Appendix II, Table 15. Figure 3 and Table 8 illustrate the first-pass regression probability for both groups, by condition (i.e. the logits were transformed into probabilities, which were expressed in percentages). Although the first-pass regression probability for the c-

condition appears to be rather high for the native speakers, it was not found to be reliably higher than that in any of the other conditions. Releveling to other conditions for the learner group did not reveal any further main effects either.
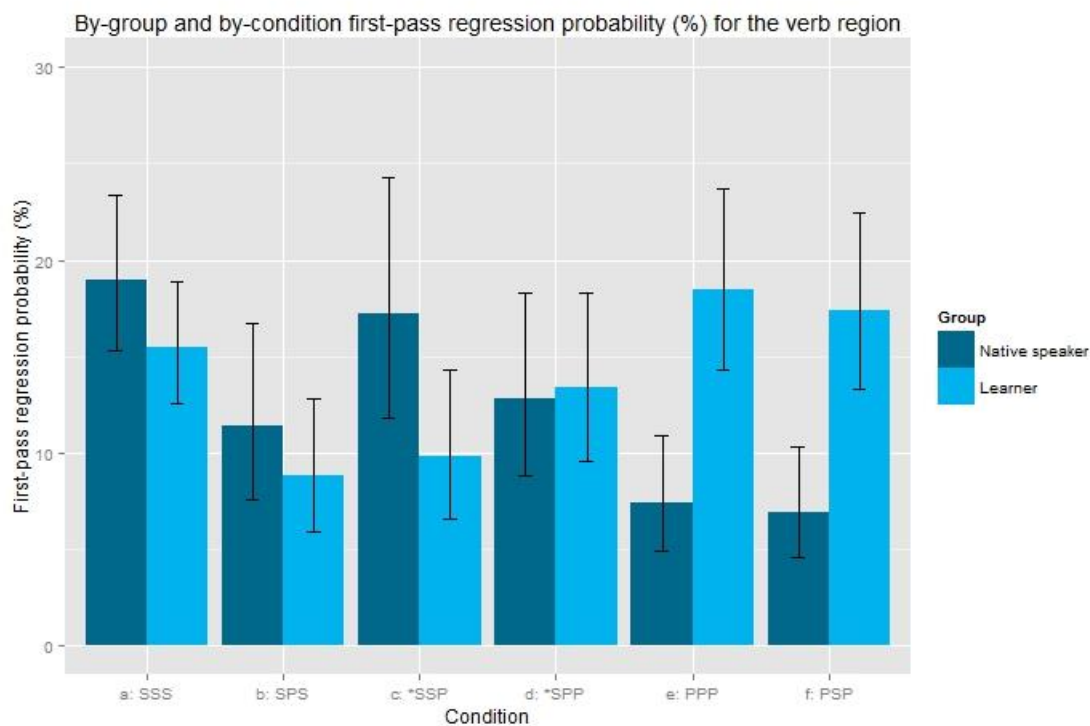
Figure 3: First-pass regression probability (%) for the verb region
The error bars represent +/- 1 standard error of the estimated mean for the factor 'Condition'.

These probabilities can also be found in Table 8.

Table 8: Probability of a first-pass regression from the verb region (%)

Group: condition	Probability (logit)	SD	Probability (%)
Native speaker: a (SSS)	-1.448	1.015	19.0
Native speaker: b (SPS)	-0.603	1.712	11.4
Native speaker: c (*SSP)	-0.125	1.692	17.2
Native speaker: d (*SPP)	-0.469	1.623	12.8
Native speaker: e (PPP)	-1.081	1.654	7.4
Native speaker: f (PSP)	-1.148	1.689	6.9
Learner: a (SSS)	-1.693	1.014	15.5
Learner: b (SPS)	-0.649	1.812	8.8
Learner: c (*SSP)	-0.527	1.807	9.8
Learner: d (*SPP)	-0.175	1.583	13.4
Learner: e (PPP)	0.211	1.319	18.5
Learner: f (PSP)	0.138	1.341	17.4

Endpoints of first-pass regressions from the verb region. The total number of first-pass regressions launched from the verb region, irrespective of group and condition, was 299. Again,

only the trials in which neither of the critical nouns was skipped during first-pass reading were considered (i.e. the original number of regression-ended first passes in the verb region was likely somewhat higher). 125 of these first-pass regressions belonged to the native speaker data and 174 to the advanced learner data. The breakdown of the regression endpoints into group, condition, and landing site can be found in Table 9 and in Figure 4 and Figure 5.

As indicated by the table and the graphs, high percentages of the first-pass regressions that were launched in the verb region landed on the word directly preceding the verb region (i.e. the local noun). Less adjacent words were targeted (much) less often, although the head noun was apparently regressed to more often than the three function words preceding the verb region, especially by the native speakers (the fact that the learners regressed to the first determiner relatively often may have caused this apparent difference). This finding is not surprising in the sense that the head noun was the only highly informative content word out of the four words preceding the local noun (and in most cases it was also the longest word). However, the percentages presented below should be interpreted with caution, since the total number of first-pass regressions that was launched by each group in each of the conditions was small (e.g. for the British group, there were only 11 and 10 observations for the e and f conditions, respectively). Therefore, statistical significance was not calculated and reported for this measure. In general, while the head noun was the most relevant landing site for the resolution of processing difficulties caused by agreement anomalies (i.e. since its number marking decides the number marking of the complex subject as a whole), the head noun was seemingly not regressed to strikingly often in ungrammatical conditions compared to grammatical ones.

Group: condition	1: determiner	2: head noun	3: preposition	4: determiner	5: local noun	Total
Native speaker: a (SSS)	2 (7.4%)	5 (18.5%)	3 (11.1%)	3 (11.1%)	14 (51.9%)	27 (100%)
Native speaker: b (SPS)	0 (0.0%)	3 (13.6%)	1 (4.5%)	4 (18.2%)	14 (63.6%)	22 (100%)
Native speaker: c (*SSP)	1 (3.2%)	12 (38.7%)	3 (9.7%)	1 (3.2%)	14 (45.2%)	31 (100%)
Native speaker: d (*SPP)	0 (0.0%)	8 (33.3%)	2 (8.3%)	0 (0.0%)	14 (58.3%)	24 (100%)
Native speaker: e (PPP)	0 (0.0%)	4 (36.4%)	1 (9.1%)	1 (9.1%)	5 (45.5%)	11 (100%)
Native speaker: f (PSP)	0 (0.0%)	5 (50.0%)	1 (10.0%)	2 (20.0%)	2 (20.0%)	10 (100%)
Learner: a (SSS)	6 (21.4%)	6 (21.4%)	2 (7.1%)	3 (10.7%)	11 (39.3%)	28 (100%)
Learner: b (SPS)	1 (4.5%)	3 (13.6%)	4 (18.2%)	3 (13.6%)	11 (50.0%)	22 (100%)
Learner: c (*SSP)	2 (8.0%)	6 (24.0%)	3 (12.0%)	1 (4.0%)	13 (52.0%)	25 (100%)
Learner: d (*SPP)	4 (12.9%)	6 (19.4%)	3 (9.7%)	0 (0.0%)	18 (58.1%)	31 (100%)
Learner: e (PPP)	4 (11.1%)	8 (22.2%)	1 (2.8%)	3 (8.3%)	20 (55.6%)	36 (100%)
Learner: f (PSP)	5 (15.6%)	7 (21.9%)	2 (6.3%)	6 (18.8%)	12 (37.5%)	32 (100%)

To illustrate with an example of an experimental sentence: 1: the 2: editor 3: of 4: the 5: newspaper *was admired* by the staff… First-pass regressions were launched from the verb region (italicised).
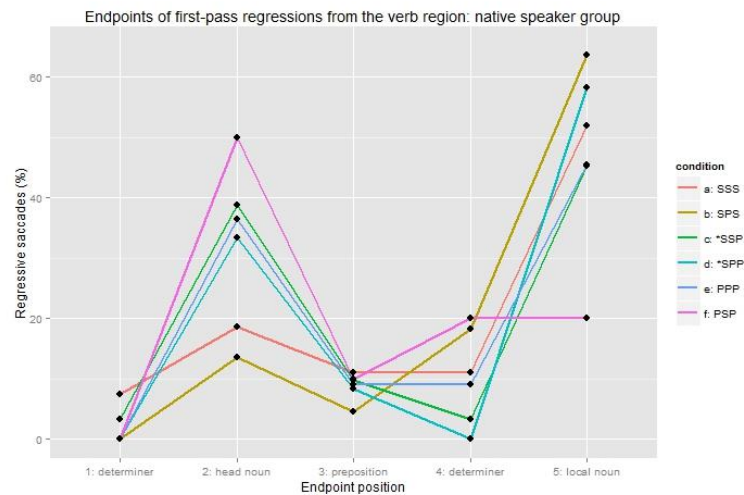

Figure 4: First-pass regression endpoints for the native speaker group
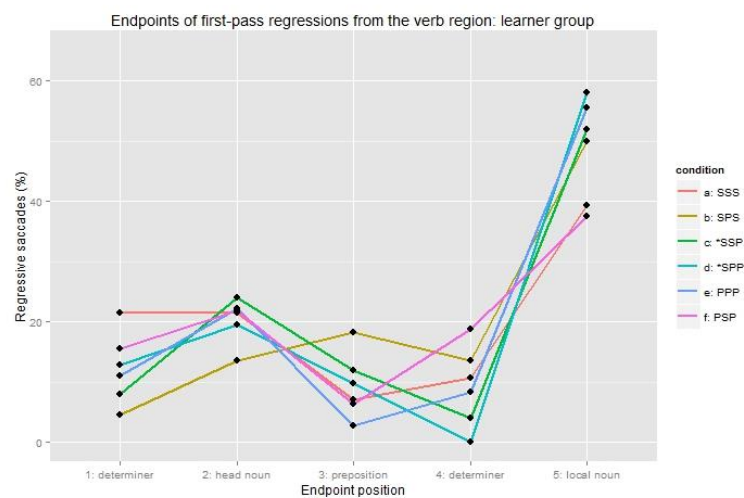
Figure 5: First-pass regression endpoints for the learner group

5.2.2 Copula

Linear mixed-effects regressions were performed on the first-pass and total reading times for the copula verb and a logistic mixed-effects regression was applied to the first-pass copula-skipping data.

First-pass reading time. No main effects were found for the first-pass reading times for the copula for either of the groups (in any of the models applied to the log-transformed and to the length-corrected log-transformed data). Moreover, adding age and the length of the word before the copula as fixed effects did not significantly improve model fit; therefore they were not included. There were no interactions between group and condition in the first-pass data. Both models with and without interaction term and with and without random slopes were applied to the data (i.e. no interactions were found and the copula skipping resulted in the loss of many data points – 40.2% – so for some participants data were very meagre or even absent for specific conditions). A subset of 26 participants who had at least two data points for each condition was then created (the 7 participants who were excluded based on this criterion were all native speakers). The analyses were repeated on this subset, but no further main effects or interactions were revealed.

Total reading time. The results for the total reading times for the copula were more similar to those found for the verb region as a whole. When the linear mixed-effects model was applied to the log-transformed reading times, main effects were found for the native speakers for the c-condition ($p < .01$) and for the learners for the d-condition ($p < .05$). When the data were further transformed into residual reading times, main effects were found for the native speakers for conditions b ($p < .05$) and c ($p < .01$). In both models, the c-condition was found to be more difficult than the a, e, and f conditions, but not than conditions b and d. An interaction was found for condition c, indicating that the learners spent comparatively little time reading the copula in this condition. No main effects were found for the advanced learners.

Probability of skipping the copula first-pass. The probability of skipping the copula first-pass was investigated using a logit mixed-effects model. Logits (Table 10) were transformed into probabilities (%), which are more intuitive and easier to interpret. The probabilities are presented in Figure 6 and Table 10. The (logit-based) regression output can be found in Appendix II, Table 16. Whereas no main effect was found for the native speakers for condition b, main effects were

found for all other conditions ($p < .001$), such that skipping was less likely for the plural-copula conditions. The same applied to the learner group; no main effect was found for condition b, but main effects were found for conditions c ($p < .001$), d ($p < .001$), e ($p < .05$), and f ($p < .001$). The four plural copula conditions did not differ significantly from each other for the native speakers, nor for the Swedish learners. No interactions were found. The main effect for group was significant, indicating that it was more likely for the native speakers to skip the copula first-pass. The e-condition was the only condition for which the between-group difference was found to be borderline significant (i.e. when the speaker group was chosen as the baseline, the difference was significant; but when the baseline was releveled to the learner group, the *p*-value was just above 0.05).
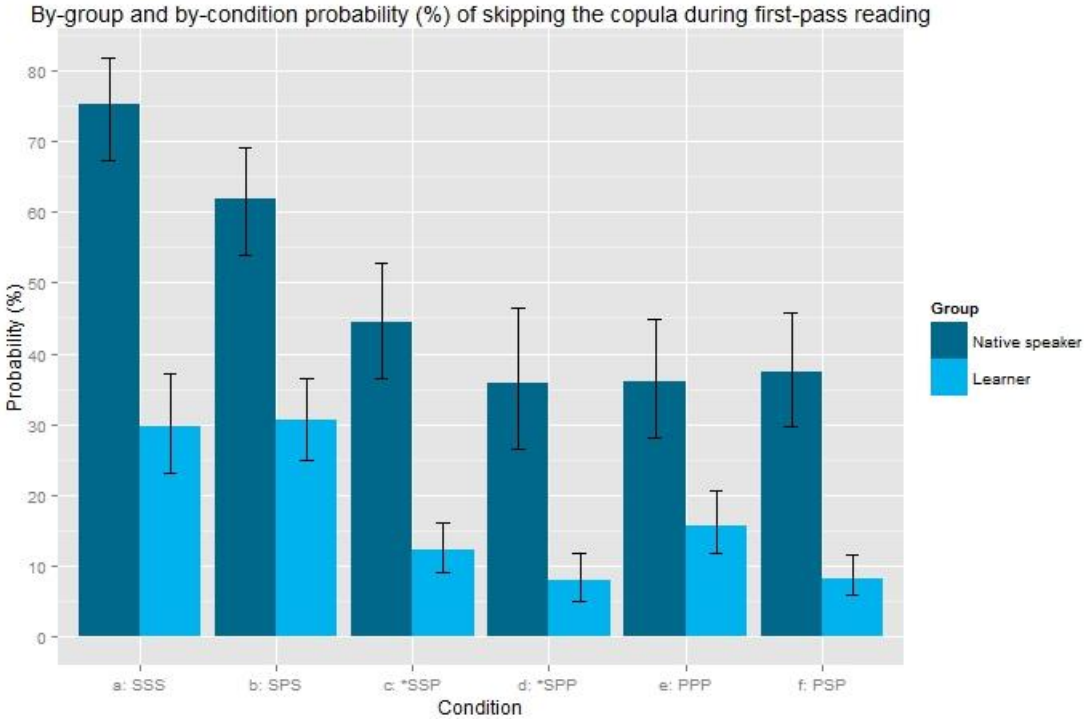
Figure 6: Probability (%) of skipping the copula during first-pass reading
The error bars represent +/- 1 standard error of the estimated mean for the factor 'Condition'.

Figure 6 and Table 11 illustrate the first-pass copula skipping probabilities. For both groups, the skipping probabilities were highest for the a and b conditions (i.e. the two singular-copula conditions). Skipping the copula first-pass in the other four conditions was approximately equally likely both within the native speaker group and within the learner group. However, as can be deduced from Figure 6 and Table 11, the between-group differences were considerable.

Table 10: Probability (logit) of skipping the copula first-pass

Group: condition	Probability (logit)	SD
Native speaker: a (SSS)	1.107	1.510
Native speaker: b (SPS)	-0.625	1.266
Native speaker: c (*SSP)	-1.333	1.294
Native speaker: d (*SPP)	-1.689	1.692
Native speaker: e (PPP)	-1.679	1.421
Native speaker: f (PSP)	-1.623	1.340
Learner: a (SSS)	-0.864	1.443
Learner: b (SPS)	0.040	1.171
Learner: c (*SSP)	-1.111	1.362
Learner: d (*SPP)	-1.597	1.930
Learner: e (PPP)	-0.816	1.404
Learner: f (PSP)	-1.534	1.540

Table 11: Probability (%) of skipping the copula first-pass and between-group differences

Condition	Probability (%) native speakers	Probability (%) learners	Difference (percentage points)	Difference (percent)
a: SSS	75.2	29.7	-45.5	-60.5
b: SPS	61.8	30.5	-31.3	-50.6
c: *SSP	44.4	12.2	-32.2	-72.5
d: *SPP	35.8	7.9	-27.9	-77.9
e: PPP	36.1	15.7	-20.4	-56.5
f: PSP	37.4	8.3	-29.1	-77.8

5.2.3   Past-participle verb

First-pass reading time. For this region of interest, 13.5% of the observations were excluded due to first-pass skipping, and an additional 0.5% of data points were excluded through data trimming. No main effects were found for the native speakers or for the advanced learners for any of the conditions. However, main effects were found for group, such that the learners spent more time in this region during first-pass reading; and for age, such that older participants were faster. No interactions were found for this measure in this region of interest.

Total reading time. A main effect was found for the native speakers for condition f ($p <$ .05), such that total reading times for this condition were reliably shorter than those for the a-condition. No main effects were found for the learner group for any of the conditions. No significant interactions were found.

5.2.4 Head noun and local noun

Total reading time. For the head noun and the local noun, respectively, 7.8% and 7.9% of all observations were excluded because of first-pass skipping, and 0.7% and 0.9% of all data points through data trimming. For the head noun, no main effects were found for either of the groups, nor were any interactions present in the data. For the local noun, no main effects were found for the native speakers, but main effects were found for the learners for conditions d ($p < .01$) and e ($p < .01$), such that the total reading times for the local noun were relatively long in these conditions. Interactions for these conditions were also found ($p < .01$). However, it should be noted that these effects could not straightforwardly be attributed to rereading alone, since main effects for these conditions ($p < .01$) – as well as for the b-condition ($p < .05$) – already appeared in the first-pass data (i.e. before the verb region or any anomaly had been encountered at all). An interaction for the d-condition ($p < .01$) was also already present in the first-pass data.

6 Discussion

The purpose of this study was to examine whether Swedish advanced L2 learners of English process subject–verb agreement, not instantiated in their L1, in target-like ways (i.e. whether their eye-movement record resembles that of native speakers). Most importantly, this study aimed to determine whether advanced learners, like native speakers, show sensitivity to violations of complex subject–verb agreement during online sentence comprehension. A secondary goal was to ascertain whether there are more general processing differences between native speakers and advanced learners.

The results indicated that, while native English speakers were sensitive to manipulations of complex subject–verb agreement online, with first-pass and total reading times varying depending on head/local noun match and grammaticality, Swedish advanced L2 learners English showed no such sensitivity. The reading times for each of the conditions were compared to those for the singular head noun, head/local noun match, grammatical baseline condition (a/SSS: the **editor** of the **newspaper was**). This condition was expected to be read and processed faster than the other conditions because it lacked any explicit grammatical number marking and because of its total grammatical number match (additionally, this condition was generally semantically most plausible). While a robust effect was found for the native speakers for the singular head noun, head/local noun match, ungrammatical condition (c/*SSP: *the **editor** of the **newspaper were**), as well as a rather robust effect for the singular head, head/local noun mismatch, grammatical condition (b/SPS: the **editor** of the **newspapers was**), no such effects were found for the learners. More general processing differences were uncovered through more detailed analyses of the copula verbs in the construction tested. These analyses revealed that native speakers were more likely to skip copulas during reading than L2 readers, suggesting a greater predictive capacity and a subtle timing advantage in processing.

6.1 Previous and present findings for native speakers

Before explicitly comparing and discussing the results for the two groups in the present study, the results for the native speakers will first be compared to those of Pearlmutter et al. (1999). It should be noted, however, that a direct comparison is complicated by the fact that the said study investigated the first four conditions – including the two ungrammatical conditions (a: SSS, b:

SPS, c: *SSP, and d: *SPP) – and all grammatical conditions – including the two plural head noun conditions (a: SSS, b: SPS, e: PPP, and f: PSP) – in three different experiments (see section 4.3.1, Table 3 for the meanings of these abbreviations). These three experiments were based on two different methodologies, i.e. self-paced reading and eye-tracking. Whereas the a, b, c, and d conditions were investigated in both paradigms, conditions e and f were only tested in a self-paced reading paradigm (alongside conditions a and b). Therefore, total reading time was the only measure available for the e and f conditions in Pearlmutter et al. (1999). Moreover, the said study employed different statistical models (i.e. 2×2 ANOVAs vs generalised mixed effects models used in the present study). Whereas Pearlmutter et al. reported 'marginally significant' main effects and interactions on several occasions, these instances will be disregarded in this discussion. Only statistically reliable main effects and interactions will be considered here (i.e. for optimal comparability, since the present study strictly adhered to a $p < 0.05$ cut-off point).

In Pearlmutter et al. (1999), the first-pass reading times for the verb, the word following the verb, and the verb region did not yield any significant main effects or interactions. However, when the regression-ended first-passes were removed (see section 5.2.1), a main effect of grammaticality was found for the verb region, alongside an interaction between grammaticality and head/local noun number match. That is, whereas number mismatches between the head noun and the local noun were found to inflate first-pass reading times in grammatical conditions, they were found to deflate first-pass reading times in ungrammatical conditions. Finally, the a-condition was found to be reliably faster than the b, c, and d conditions.

The present study replicated these results, insofar as no main effects were found for first-pass reading times for the verb or for the past-participle verb. A main effect was found for the c-condition in the verb region even before regression-ended cases were removed. After the removal of these observations, main effects were found for the b and c conditions, but not for the d condition, i.e. condition a was read faster than conditions b and c, but not d. Although interactions between within-subjects factors in the present study were more implicit in nature (i.e. the conditions – combinations of different levels of each factor of interest – were entered into the statistical models as a whole), it can still be concluded from these results that head noun/local noun number mismatches indeed increased processing difficulty in grammatical conditions, while they deflated first-pass reading times in ungrammatical conditions. However, while Pearlmutter

et al. stated that the effects of ungrammaticality were stronger than those of mismatches in head noun and local noun number, the present study found that the latter eased the processing in the ungrammatical d condition to the extent that no difference was found between this condition and the baseline condition.

In terms of total reading times, Pearlmutter et al. found a grammaticality and head/local noun NP-match interaction at the copula, as well as main effects of grammaticality and NP-match. Compared to the a condition, total reading times were found to be longer for conditions c and d. Condition c was found to be more difficult than any of the other conditions. At the word following the verb, a main effect of grammaticality was found, as was an interaction between grammaticality and NP-match. Condition c was reliably more difficult than the other three conditions, and the a-condition was reliably easier than the other three conditions.

The total reading time results in the present study did not closely resemble those in Pearlmutter et al. (1999). That is, at the copula, main effects were found for conditions b and c (i.e. reading times were reliably longer in these conditions), not condition d. Moreover, condition c was not found to be more difficult than any condition other than condition a (and conditions e and f). For the word following the verb, the present study found no reliable main effects other than for condition f. While the 1999 study did not investigate total reading times for the verb region as a whole, the present study did. A significant main effect was found for condition c, which was also found to be reliably more difficult than any of the other conditions.

Pearlmutter et al. concluded from their total reading time results that, over time, the difference between the grammatical a and b conditions (as reflected in the first-pass reading times) decreased, while the relative difficulty of the ungrammatical c and d conditions increased (in particular that of the c-condition). They argued that, overall, "NP-mismatch tended to be less disruptive than ungrammaticality" (Pearlmutter et al., 1999, p. 443). On the contrary, the present study found that the effect of a head/local noun number mismatch in the grammatical b-condition was still visible in total reading times at the copula, whereas no main effects were found for the mismatch, ungrammatical d-condition in any of the regions of interest. These results indicate that the effect of a head/local noun number mismatch may substantially diminish the processing difficulty associated with ungrammaticality, even in late processing. It should be noted that the effect that was found for the b-condition may to some extent also be related to semantic

processing difficulties (i.e. for several of the experimental sentences, the b-condition was semantically less plausible than the other conditions). In line with the 1999 study, the c-condition did clearly induce the most severe processing difficulty.

Additionally, Pearlmutter et al. investigated total reading times for the head noun and the local noun, which were found to be longer for the ungrammatical conditions compared to the grammatical ones. However, the present study revealed no such pattern; no main effects were found for any of the conditions. In terms of first-pass regression probabilities, Pearlmutter et al. found reliably higher regression rates for the c-condition compared to the others, as well as main effects for and an interaction between grammaticality and NP-match. This finding was taken as evidence that higher regression probabilities may indeed be indicative of processing difficulty. However, the present study did not find any main effects for first-pass regression probability other than those for the plural head noun, grammatical e and f conditions (such that first-pass regressions were less common in these conditions). Both studies found that the endpoints of first-pass regressions launched from the verb region were not strongly targeted by syntactic processes. That is, the head noun, which carried the relevant number marking, was regressed to much less often than the more adjacent local noun. Moreover, no clear differences were found between grammatical and ungrammatical conditions at the head noun position. Pearlmutter et al. argued that the readers in their study eventually did target the appropriate locations for reanalysis, based on their finding that the total reading times for the head noun and the local noun reflected differences between the grammatical and ungrammatical conditions. The present study did not replicate this finding, however. Moreover, their finding that the head noun and local noun received relatively many passes compared to the two determiners and the preposition preceding the verb region was discussed as potentially meaningful. However, determiners and prepositions are typically short function words that are skipped relatively often. Therefore, this argument does not carry much weight.

Pearlmutter et al. compared the two plural head noun conditions (e and f) and two of the singular head noun conditions (a and b) in a self-paced reading experiment. All conditions adopted in this experiment were grammatical. 2 × 2 ANOVAs with factors head noun number and head/local noun number match were applied to the total reading time data. At the copula verb, a main effect for head/local noun number match was found, alongside an interaction. While

the verb was processed reliably faster in condition a compared to condition b, there were no differences between the two plural head noun conditions. On the word following the verb, however, the difference between the a and b conditions was no longer present, whereas a rather surprising reverse effect was found for the plural head conditions. That is, the mismatch condition (f) was found to be easier than the match condition (e).

Since the present study included all conditions in a single eye-tracking experiment, several measures were available for the e and f conditions. A main effect was found for the f-condition for total reading time at the past-participle verb, such that this region was read relatively fast compared to the baseline condition. No further main effects were found for either of these conditions for the time-based measures. When directly comparing the two conditions to each other, small numerical differences in the direction described by Pearlmutter et al. were found, but none of them reached significance (i.e. this also applied to the total reading time at the past-participle verb). So, the reverse mismatch effect for the plural head noun conditions found in Pearlmutter et al. (1999) was not replicated. Finally, first-pass regressions were less likely launched from the verb region in both of these conditions compared to the baseline condition. Since both plural head noun conditions were grammatical and since plural subjects are usually not subject to any form of agreement attraction, these findings are not highly surprising.

Within the framework of research on agreement attraction phenomena, the results of the present study can both be explained from the perspective of percolation theories as well as from that of theories which assume that these phenomena emerge when subject number is re-assessed at the verb (see section 2.1.1). That is, percolation of number features up to the head noun position can explain the relative difficulty of the grammatical b-condition, as well as the absence of an ungrammaticality effect for the d-condition. Moreover, attraction effects are typically not observed for plural subjects. Since the local noun in the f-condition is not marked for grammatical number (i.e. it is singular), no feature is available for percolation up to the head noun. The findings of the present study also fit into the second type of account mentioned above. In cases where the head noun is singular, the local noun that is explicitly marked for grammatical number is accessed and processed (partly or completely) just before the verb is first accessed. Simply put, in such accounts the copula verb is thought to give off cues of singularity (b-condition) and plurality (d-condition), among others. Upon re-accessing information about the

subject at the verb, the number of the head and local noun may be confused, explaining the seeming ungrammaticality of the b-condition as well as the seeming grammaticality of the d-condition. The finding that the c-condition was more difficult than any of the other conditions can also be explained from both perspectives. That is, no number feature is available for percolation in this condition, nor can the number of the two nouns be confused at the verb, since it matches. The mismatch in subject and verb number is therefore rather unmistakable.

Although the results of the present study do not favour one set of theories in particular, Wagers et al.'s (2009) notable finding that attractor effects may occur even when the attractor noun does not intervene directly between the subject and the verb (and when it is hierarchically superior) should be taken into account. Interestingly, Wagers et al. did not find an attractor effect for grammatical sentences in their self-paced reading experiments, which was taken as additional evidence against theories that attribute attraction phenomena to misrepresentations of subject number. However, the present study did find the type of symmetry in grammatical and ungrammatical conditions predicted by this type of account.

In sum, focusing on native speakers, the results from the present study were found to diverge from those in Pearlmutter et al. in some important respects. It is possible that these differences were partly caused by the statistical modelling. It could be that fixed and random effects in the generalised mixed-effects models used in the present study accounted for variance that was not captured by the ANOVAs applied in the said study. Moreover, under the surface, the designs of the studies diverged in several ways (see especially section 4.3.2). It should also be noted that the number of items used for the said study was low (4 of each type and condition), and because of the removal of trials (e.g. necessitated by skipping, the software failing to detect the pupil or the corneal reflection, or inaccurate answers to accompanying comprehension questions), there was probably some unbalance in the usable numbers of trials of each type and condition, especially affecting the data for the eye-tracking experiment. This may also have influenced the results to some extent, i.e. in terms of high variance associated with items and also because ANOVA analyses are not robust against unbalanced datasets. The numbers of participants tested in each of their experiments, on the other hand, were much higher than that in the present study.

6.2 Advanced learners versus native speakers

While the different findings for native speakers are both relevant and interesting, the main question addressed by the present study was whether Swedish advanced L2 learners of English process subject–verb agreement manipulations in English sentences in qualitatively similar ways compared to native speakers of the language. Therefore, the findings for both groups will be compared and contrasted here. First-pass reading time (or gaze duration) was expected to be the dependent measure with the strongest potential to reveal qualitative similarities or differences between the two groups (see section 4.5.2.1). When the statistical models were applied to the residual reading times, no significant main effects were found for the advanced learners for first-pass reading times at the copula, the past-participle verb, and in the verb region. While the same applied to the native speakers for the copula verb and the past-participle verb, main effects were found for this group for the b and c conditions in the verb region, such that these conditions were relatively difficult to process. An interaction for these conditions was also found, indicating that, for the learners, the divergence from the baseline condition was relatively small (i.e. while it was large enough to be reliable for the native speakers).

These findings for the first-pass reading times suggest that the seeming ungrammaticality effect of the b-condition and the actual ungrammaticality effect of the c-condition that were found for the native speakers were not present in the advanced learner data. That is, compared to the baseline condition, these conditions did not seem to induce heavier processing costs for the learners.

In this context, it should be noted that main effects for the advanced learners were found for the c and f conditions before the first-pass reading times were transformed into residual reading times. Although it appears that these effects were an artifact of region length (i.e. they disappeared when the first-pass reading times were transformed into residual reading times), it should be noted that the length-correction method that was applied was not ideal. That is, the individual regression equations were based on the relationship between the word length of the region of interest and the (first-pass) reading times for that region. Although the differences in word length between conditions were minimal (zero or one characters), it would have been preferable to base the individual regression equations on (first-pass) reading times for all words from all experimental sentences and filler items (i.e. in that case, the equations would have

probably been even more accurate). Therefore, these main effects cannot be completely disregarded. However, even though the main effect for the c-condition in the non-residualised output (Table 12) could imply online sensitivity to ungrammaticality, an interaction was found for the b-condition, indicating that the learners spent comparatively little time reading this region. Moreover, the main effect for the f-condition was such that this condition increased processing difficulty for the learners, which did not apply to the native speakers. So, even if the results for the non-residualised first-pass reading times were to be taken into account instead of those for the residualised reading times, the results for the two groups would differ. However, the region length correction methods applied to the data seem appropriate considering the small between-condition length differences and the rigidity of the structure (i.e. causing the regions of interest to have stable positions across sentences). Moreover, the data were more normally distributed after the residualisation.

The findings for the total reading times were rather similar. That is, no significant main effects were found for the advanced learners for any of the regions of interest. For the native speakers, on the other hand, main effects for the b and c conditions were found at the copula and a main effect for the f-condition was found at the past-participle verb, such that this region was read relatively fast in this condition. In the verb region, a significant main effect for the c-condition was found. Both at the copula and in the verb region, interactions between group and condition c reached significance, suggesting that the learners spent less time reading these regions than would be expected if they were similarly sensitive to the ungrammaticality online. Total reading times on the local noun were longer for the advanced learners in the d and e conditions and interactions were also found. However, since these effects were already present in the first-pass reading time data, before the verb region was read for the first time, these results cannot be considered very meaningful.

In terms of first-pass regression probability, it was not apparent from the native speakers' results that ungrammaticality led to higher rates of first-pass regressions from the verb region. In fact, the only main effects found for the native speakers were those for conditions e and f, indicating that regression probabilities from these regions were relatively low. These effects could be significant by coincidence (since this is the only measure which revealed effects for these conditions). Another possibility is that the plural marking on the head noun in these conditions

created a relatively strong imprint; thereby making it less necessary for readers to re-access properties of the subject through regression and rereading. Interactions were also found for these two conditions, indicating that the advanced learners regressed from the verb region comparatively more (see also Figure 3). For the advanced learner group, no main effects were found for this measure. However, since there were also no straightforward ungrammaticality or mismatch effects for the native speakers, this finding was not taken to be very meaningful.

Related to the regression probabilities were the endpoints of first-pass regressions launched from the verb region. For both groups, the most common landing site for a first-pass regression from this region was the directly adjacent local noun. While the head noun, which carried the relevant number marking, was regressed to more often than the three function words preceding the verb region (especially by the native speakers), there was no clear evidence of syntactically driven targeting. That is, the differences between grammatical and ungrammatical conditions were generally small and not particularly straightforward. While the native speakers rarely regressed back to the first determiner, it was a relatively common final landing site for first-pass regressions launched by the advanced learners. Possibly, the native speakers more effectively targeted the most informative word, while learners more commonly reread the first part of the sentences in full. It should be noted once more that the number of regression-ended passes for each group and for each condition was too small to be able to draw any sound conclusions based on this measure.

The probability of skipping the copula first-pass was investigated to establish whether there were more general processing differences between the two groups. Within the groups, the skipping patterns were highly similar. The singular copulas in conditions a and b were skipped relatively often compared to the one character longer plural copulas in conditions c, d, e, and f. Within the groups, there were no differences between the a and b conditions or between the c, d, e, and f conditions. Between the groups, the differences were considerable and significant for each condition. As an illustration, while the copula-skipping probabilities for the native speakers for the a and b conditions were 75.2% and 61.8%, respectively, those for the advanced learners were 29.7% and 30.5%, respectively. In percentages, the between-group differences amounted to 60.5% for the condition a and 50.6% for condition b. The learners were always at least 50.6% (condition b) and at most 77.9% (condition d) less likely to skip the copula during first-pass

reading.

It is likely that these considerable differences in copula-skipping probabilities were caused by differences in reading behaviour and strategies between the native speaker group and the advanced learner group. That is, it may be that the advanced learners were more motivated to (or felt more pressure to) read the sentences carefully because they were performing in their second language (in a test setting). It may also be that they needed to read the sentences relatively carefully from the point of view of comprehension. For the native speakers, on the other hand, it may often have sufficed to mainly fixate content words for comprehension purposes. In line with this, it could be that the native speakers were subject to more general processing benefits, i.e. that they were more successful at predicting upcoming information and at processing information parafoveally (while learners likely needed more time to access and to process words).

To summarise, in addition to more general processing differences as reflected in the copula-skipping probabilities, it appears that the advanced learners in the present study did not process the subject–verb manipulations in the present study in qualitatively similar ways to the native speakers. That is, the learner data did not reflect online sensitivity to real or seeming violations of complex subject–verb agreement, while these increased processing difficulty for the native speakers. In this context, it is also interesting to note that the post-experiment interviews revealed that only one out of the full set of 22 Swedish participants noticed any grammatical number mistakes in the sentences (this person did not realise these were included intentionally).

The apparent lack of online sensitivity could potentially be explained by limitations in L2 proficiency and factors such as working memory capacity. Especially when non-local dependencies are handled, working memory capacity is highly relevant. Since processing in the L2 consumes more cognitive resources compared to processing in the L1, working memory capacity may be reduced for L2 learners compared to native speakers. However, it has been found that less working memory capacity is consumed when L2 learners become more experienced/proficient (e.g. Service et al., 2002). Participants in the present study were required to have an advanced English proficiency level. Their high reading comprehension, LexTALE, and multiple choice accuracies indicated that this requirement was generally met. Moreover, the LHQ 2.0 revealed that all Swedish participants engaged in activities involving English on a daily basis. Excluding the two participants with the lowest LexTALE scores (statistical outliers) did not

change the pattern of main effects in any meaningful way, nor was the LexTALE score a reliable predictor in the statistical models applied to the first-pass and total reading times. More detailed analyses of the reading time patterns for the five highest achievers in terms of L2 proficiency (these participants scored high across the board; not only on the LexTALE task) did not reveal any clear effects for number mismatches or ungrammaticality either.

The L2 learners' reading speed was lower than that of the native speakers and it is conceivable that this may obscure effects. However, processing speed differences between L2 learners and native speakers (i.e. differences in time needed for e.g. lexical, syntactic, and semantic access) are typically found (Fender, 2001) but do not seem "to qualitatively affect bilinguals' processing procedures" (Roberts, 2013b, p. 238). That is, online sensitivity to specific (agreement) manipulations may be found despite the generally longer reading times for L2 readers (e.g. Foucart & Frenck-Mestre, 2012; Keating, 2009, Keating, 2010). Moreover, the typological distance between the learners' L1 and L2 in the present study was minimal compared to that in previous studies on related dependencies. In fact, Swedish even instantiates certain types of number agreement, as opposed to the L1s in these studies. Furthermore, the eye-tracking method was chosen because of its relatively strong potential to reveal effects for non-local dependencies. That is, methods applied in previous studies on related dependencies, such as self-paced reading and ERP, arguably put a heavier strain on working memory resources (thereby potentially leading to comparatively weak signal-to-noise ratios).

Therefore, the fact that no online sensitivity was found for the learners is better explained in terms of the structural distance involved and/or the complexity of the structure investigated in the present study. The subject NP has an internal structure, as its head noun is modified by a PP. This PP, consisting of a preposition and an NP, creates structural distance between the (dis)agreeing elements. The SSH prediction for this type of non-local dependency is that target-like processing should not be possible for L2 learners (Clahsen & Felser, 2006 a,b,c). Since previous research has indicated that qualitatively similar processing may in fact be possible for specific syntactic structures involving non-local dependencies (e.g. Foucart & Frenck-Mestre, 2012), it could be that this particular structure is too complex for advanced learners to be able to achieve target-like processing. That is, the grammatical number of the head noun needs to be retained in working memory and the subject as a whole needs to be substituted to be able to

accurately conflate person and number at the verb. Both linear and structural distance between the (dis)agreeing elements and the presence of a distracting attractor noun in the modifying PP complicate this.

In terms of implications for the SSH (Clahsen & Felser, 2006 a,b,c), the results of the present study appear to provide support for its predictions. However, the structure investigated does involve a rather large linear distance (of three words) alongside a smaller structural distance, and the design does not allow for their effects to be teased apart. Moreover, since target-like processing was found to be possible for certain non-local dependencies (see section 2.2.2), the hypothesis may need some refinement. It is indeed possible that target-like processing of complex subject–verb agreement is not achievable for L2 learners. However, further research is needed to establish whether this applies to L2 learners in general, or more specifically to L2 learners whose L1 does not instantiate this type of agreement (cf. Hopp, 2010). To resolve this issue, investigating another learner group from an L1 background that does instantiate subject–verb agreement would be a potential solution (this L1 background should preferably not be typologically remote). In fact, more potentially relevant factors need to be investigated to arrive at more conclusive answers, including individual differences in proficiency (cf. Hopp, 2010) and working memory capacity (cf. Keating, 2010).

As regards the Full Transfer / Full Access Hypothesis (Schwartz & Sprouse, 1994, 1996) and the Competition Model (MacWhinney, 1987, 2005), it should be noted that the absence of online sensitivity for the advanced L2 learners in the present study per se cannot be considered as evidence against the models' predictions. That is, as discussed, the (linear and) structural distance and/or the structural complexity of the configuration investigated may have prevented target-like processing of this particular syntactic structure unique to the L2. Using a similar L2 learner group to investigate L2 processing of simple subject–verb agreement alongside complex subject–verb agreement could potentially lead to more compelling insights into these matters.

A final plausible explanation for the absence of online sensitivity in the learner data lies in the fact that the number of advanced learners tested in the present study was fairly small ($n = 18$). There may have been rather large individual differences between participants, which may have resulted in noise that obscured or cancelled out effects in the data. A larger sample of participants could generate a more robust data set and lead to more conclusive answers. Individual differences

in for example proficiency level, age, and reading pace were taken into account as much as possible, but additional factors may have been of importance.

Another limitation to the present study was that the participants' offline knowledge of the workings of complex subject–verb agreement (as well as their online sensitivity to violations of simple subject–verb agreement) was not independently attested. Therefore, no direct claims can be made about (a)symmetry between offline and online results. Nonetheless, the lack of online sensitivity is fairly surprising considering the fact that participants generally scored highly on the independent tests. Although the LexTALE is a lexically rather than a syntactically based task, it has been found to give a fair indication of general proficiency. The lack of a subject–verb agreement system in L1 Swedish probably complicates the acquisition of the rule in L2 English, but it is not likely the primary reason for the absence of online sensitivity in the present study. That is, several studies have found that syntactic structures not instantiated in the L1 may be processed in target-like ways at higher L2 proficiency levels. It is possible that the learners tested in the present study did not have fully internalised procedural knowledge of the subject–verb agreement rule in their L2. Although the production of subject–verb agreement was not investigated for these specific L2 learners, it has been attested that subject–verb agreement may cause persistent problems in production for Swedish L2 learners of English, even at higher proficiency levels. In this case, comprehension may mirror production, with only L2 learners with English proficiency levels surpassing a certain threshold (e.g. near-native) being able to process and produce the structure in target-like ways. Moreover, as discussed, subject–verb agreement in English is not often obviously required from the point of view of sentence comprehension. Nonetheless, it has been found that native speakers process it automatically when reading sentences for comprehension. L2 learners, on the other hand, may adopt different reading strategies when faced with a sentence comprehension task (i.e. slower and more careful reading). In the present study, any effects of mismatches and/or ungrammaticalities may therefore have been attenuated to the extent that they cancelled out altogether. It would be interesting to investigate whether the online results would be different using a grammaticality judgement paradigm instead of a comprehension paradigm. While the L2 learners in the present study probably allocated much attention (processing capacity) to meaning over form, the grammaticality judgement paradigm would have an opposite effect.

7 Conclusion

Consistent with previous research and in accord with the main hypothesis, this reading study has demonstrated that Swedish advanced L2 learners of English do not process complex subject–verb agreement manipulations in English in qualitatively similar ways to native speakers. The advanced learners' first-pass and total reading times did not reveal online sensitivity to real and seeming violations of complex subject–verb agreement, as opposed to those of the native speakers.

Several possible explanations were discussed. Three important considerations were meant to overcome issues with previous research on the processing of agreement relations unique to the L2. Firstly, the present study looked at a more strongly typologically related L1–L2 pairing compared to most of the previous studies of similar structures. Swedish, the L1 of the learners in the present study, even instantiates certain types of number agreement, as opposed to for example Chinese and Japanese. Moreover, the current study adopted eye-tracking as a method. This method is highly suitable for investigating non-local dependencies, since it does not put as much of a strain on working memory resources compared to methods that have typically been applied to investigate non-local agreement relations, such as SPR and ERP. Another advantage was this method's fine grain of analysis, allowing for the comparison of several early and late dependent measures. Thirdly, the strong word-level control in this study (i.e. in terms of lexical frequency, the exclusion of cognates and homographs, etc.) was essentially meant to prevent issues with lexical access, since this may be much faster in case of cognates and slower in case of low-frequency words. Despite these methodological improvements, no online sensitivity was found for any of the manipulations in the advanced learners' eye-movement records. Importantly, this finding implies that the absence of online sensitivity in previous studies on the processing of non-local agreement relations was likely not a direct by-product of typological remoteness, and also that it was likely not primarily method-induced.

Since the learners' L2 proficiency level was controlled for and since there was no clear correlation between individual differences in proficiency and the absence or presence of effects, the structural distance involved in and the structural complexity of the construction under scrutiny were considered more likely explanations for the absence of main effects for real and seeming agreement violations in the learner data. In terms of implications for the SSH (Clahsen & Felser, 2006 a,b,c), the evidence converges with the predictions of the hypothesis. However,

since previous research has indicated that target-like processing of certain types of non-local agreement relations may in fact be possible under certain circumstances, the hypothesis may need some refinement. That is, further research is needed to establish which are the types of non-local syntactic dependencies for which target-like ('deep') processing are and are not possible, and why (e.g. when does structural distance become too large, does this differ depending on the type of agreement relation, which types of intervening phrases or word classes are likely to prevent target-like processing and because of which qualities?, etc.). The groups of L2 learners to which these predictions apply should also be further specified. That is, the L1-background of the learners and the possibility of L1-transfer should be taken into account, alongside other relevant factors such as individual differences in proficiency and working memory capacity. Moreover, future research should attempt to further tease apart the effects of linear and structural distance on L2 processing (cf. Alemán Bañón et al., 2014; Keating, 2010).

Apart from the qualitative between-group differences in agreement processing, more general processing differences were attested when the first-pass copula-skipping probabilities were investigated. Within both groups, the skipping patterns were found to be highly similar. However, the between-group differences for each of the conditions were considerable, i.e. the learners were always at least 50% less likely to skip the copula during first-pass reading. This finding has interesting implications for theories of how reading and processing in the L2 may generally differ from reading and processing in the native language (i.e. at least in test settings). Similar analyses of other function words in the experimental sentences used in the present study (e.g. determiners and prepositions) could potentially lead to more conclusive insights. Because of the rigidity of the complex subject–verb agreement configuration, the positions of the subject-internal determiners and preposition are stable across sentences, allowing for straightforward comparisons.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition, 30*(4), 481–509.

Alemán Bañón, J., Fiorentino, R. and Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1–L2 similarity and structural distance. *Second Language Research, 30*(3), 275–306.

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A, 58*(6), 1065–1086.

Balota, D.A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17*(3), 364–390.

Barr, D. J., Levy R., Scheepers C., Tily H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D., Maechler, M., Bolker, B.M., and Walker, S. (2014). "lme4: Linear mixed-effects models using Eigen and S4". R package, version 1.1-7, http://CRAN.R-project.org/package=lme4.

Bock, K., & Miller, C.A. (1991). Broken Agreement. *Cognitive Psychology, 23*(1), 45–93.

Cambridge Dictionaries Online. (1995–). http://dictionary.cambridge.org/.

Capel, A. (2010). A1–B2 vocabulary: Insights and issues arising from the English Profile Wordlists project. *English Profile Journal, 1*(1), e3.

Capel, A. (2012). Completing the English Vocabulary Profile: C1 and C2 vocabulary. *English Profile Journal, 3*(1), e1.

Chen, L., Shu, H. U. A., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*(2), 161–174.

Clahsen, H., & Felser, C. (2006a). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*(1), 107–126.

Clahsen, H., & Felser, C. (2006b). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42.

Clahsen, H., & Felser, C. (2006c). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–570.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R.P.G. van Gompel, M.H. Fischer, W.S. Murray, & R.L. Hill (Eds.), *Eye movements: A window on mind and brain.* Oxford, UK: Elsevier Ltd., 341–372.

Cook, V. J., & Newson, M. (2007). *Chomsky's Universal Grammar: An introduction.* Malden, MA: Blackwell Publishing.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369–382.

Davies, M. (2008–). *The Corpus of Contemporary American English: 450 million words, 1990-present,* http://corpus.byu.edu/coca/.

De Groot, A. M. B., Borgwaldt, S., Bos, M., & Van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language, 47*(1), 91–124.

Dufour, R., & Kroll, J.F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory and Cognition*, *23*(2), 166–180.

Dussias, P.E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics*, *30*, 149–166.

Eberhard, K.M. (1997). The marked effect of number on subject–verb agreement. *Journal of Memory and Language, 36*(2), 147–164.

Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject–verb agreement in English. *Journal of Memory and Language, 41*(4), 560–578.

Eberhard, K. M., Cutting, J. C., and Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review, 112*(3), 531–559.

Ellis, N.C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 164–194.

Fender, M. (2001). A review of L1 and L2/ESL word integration skills and the nature of L2/ESL word integration development involved in lower-level text processing. *Language Learning, 51*(2), 319–396.

Ferreira, F., & Clifton, C.E. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*(3), 348–368.

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language, 66*(1), 226–248.

Franck, J., Vigliocco, G., and Nicol, J. (2002). Attraction in sentence production: The role of syntactic structure. *Language and Cognitive Processes, 17*(4), 371–404.

Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second LanguageResearch, 21*(2), 175–198.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in cognitive sciences, 6*(2), 78-84.

Gillon Dowens, M., Guo, T., Guo, J., Barber, H.A., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish – Evidence from Event Related Potentials. *Neuropsychologia, 49*(7), 1651–1659.

Gillon Dowens, M., Vergara, M., Barber, H.A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience, 22*(8), 1870–1887.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 30*(3), 251–266.

Hahne, A., Mueller, J.L., & Clahsen, H. (2006). Morphological processing in a second language: Behavioural and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience, 18*(1), 121–134.

Hawkins, R., & Chan, C.Y.H. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. *Second Language Research, 13*(3), 187–226.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures.* Oxford University Press.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua, 120*(4), 901–931.

Hoshino, N., Dussias, P.E., & Kroll, J. F. (2010). Processing subject–verb agreement in a second language depends on proficiency. *Bilingualism: Language and Cognition, 13*(2), 87–98.

Ingram, J. C. (2007). *Neurolinguistics: An introduction to spoken language processing and its disorders.* Cambridge University Press.

Inhoff, A.W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics, 40*(6), 431–439.

Jiang. N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*(04), 603–634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning, 57*(1), 1–33.

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1312–1318.

Just, M.A., & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354.

Källkvist, M., & Petersson, S. (2006). An *s* or not an *s*; that is the question: Swedish teenage learners' explicit knowledge of subject–verb agreement in English. In Einarsson, J., Larsson Ringqvist, E.K., & Lindgren, M (Eds.). *Språkforskning på didaktisk grund: Rapport från ASLA:s höstsymposium*. Växjö University.

Karlsson, M. (2002). *Progression and Regression. Aspects of Advanced Swedish Students' Competence in English Grammar*. Lund University.

Keating, G.D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning 59*(3), 503–535.

Keating, G.D. (2010). The effects of linear distance and working memory on the processing of gender agreement in Spanish. In B. van Patten, & J. Jegerski (Eds.), *Research in second language processing and parsing*. Amsterdam: John Benjamins, 113–134.

Köhlmyr, P. (2003). *"To Err is Human": An investigation of grammatical errors in Swedish 16-year-old learners' written production in English*. Gothenburg University.

Kormos, J. (2013). Sentence Production in a Second Language. In C.A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Blackwell Publishing Ltd.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325–343.

Levy, R. (2013). Memory and surprisal in comprehension. In R.P.G. van Gompel (Ed.), *Sentence processing.* Hove: Psychology Press, 78–114.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition, 17*(3), 673–680.

Liversedge, S.P., Paterson, K.B., & Pickering, M.J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception.* Oxford, UK: Elsevier Ltd., 55–75.

MacWhinney, B. (1987). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition.* Hillsdale, NJ: Lawrence Erlbaum, 249–308.

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll, & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches.* New York: Oxford University Press, 49–67.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity in sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior, 23*(2), 127–150.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition, 27*(1), 53–78.

Nicol, J. & Greth, D. (2003). Production of subject–verb agreement in Spanish as a second language. *Experimental Psychology, 50*(3), 196–203.

Ojima, S., Nakata, H., & kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17*(8), 1212–1228.

Pearlmutter, N.J., Garnsey, S.M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41*(3), 427–456.

Pollatsek, A., Rayner, K, & Balota, D.A. (1986). Inferences about eye-movement control from the perceptual span in reading. *Perception and Psychophysics, 40*(2), 123–130.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506.

Rayner, K., & Duffy, S.A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition, 14*(3), 191–201.

Rayner, K., Kambe, G., & Duffy, S.A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology, 53*(4), 1061–1080.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ: Prentice-Hall.

Rayner, K., & Pollatsek, A. (2006). Eye-movement control in reading. In M.J. Traxler, & M.A. Gernsbacher (Eds.), *Handbook of psycholinguistics.* London: Elsevier Inc., 613–657.

Rayner, K., Warren, T., Juhasz, B.J., & Liversedge, S.P. (2004). The effects of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*(6), 1290–1301.

Rayner, K., & Well, A.D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin and Review, 3*(4), 504–509.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/.

Reichle, E.D., Rayner, R., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*(4), 445–476.

Roberts, L. (2012). Review article: Psycholinguistic techniques and resources in second language acquisition research. *Second Language Research, 28*(1), 113–127.

Roberts, L. (2013a). Sentence and discourse processing in second language comprehension. In C.A. Chapelle (Ed.), *Encyclopedia of Applied Linguistics.* Blackwell Publishing Ltd.

Roberts, L. (2013b). Sentence processing in bilinguals. In R.P.G. van Gompel (Ed.), *Sentence processing.* Hove: Psychology Press, 221–246.

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition, 30*(3), 333–357.

Roberts, L., & Liszka, S.A. (2013). Processing tense/aspect-agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English. *Second Language Research, 29*(4), 413–439.

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition, 35*(2), 213–235.

Rossi, S., Gugler, M.F., Friederici, A.D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18*(12), 2030–2048.

Sabourin, L., & Stowe, L.A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research, 24*(3), 397–430.

Schotter, E.R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention Perception, & Psychophysics, 74*(1), 5–35.

Schwartz, B.D., & Sprouse, R.A. (1994).Word order and nominative case in nonnative language acquisition: A longitudinal study of (L1 Turkish) German Interlanguage. In Hoekstra, T. & Schwartz, B.D (Eds.), *Language acquisition studies in generative grammar: papers in honor of Kenneth Wexler from the 1991 GLOW workshops.* Philadelphia, PA: John Benjamins, 317–68.

Schwartz, B.D., & Sprouse, R.A. (1996). L2 cognitive states and the Full Transfer / Full Access model. *Second Language Research, 12*(1), 40–72.

Sereno, S.C., O' Donnell, P.J., & Rayner, K. (2006). Eye movements and lexical ambiguity resolution: Investigating the subordinate-bias effect. *Journal of Experimental Psychology: Human Perception and Performance, 32*(2), 335–350.

Service, E., Simola, M., Metsänheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology, 14*(3), 383–408.

Solomon, E.S., & Pearlmutter, N.J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology, 49*(1), 1–46.

Steinhauer, K., White, E., King, E., Cornell, S., Genesee, F. & White, L. (2006). The neural dynamics of second language acquisition: Evidence from event-related potentials. *Journal of Cognitive Neuroscience Supplement, 99.*

Thagg Fisher, U. (1985). *The sweet sound of concord – A study of Swedish learners' concord problems in English.* Lund studies in English, 73.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27*(2), 173–204.

Trenkic, D., Mirkovic, J., & Altmann, G.T.M. (2013). Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition, 17*(2), 237-257.

Trueswell, J.C., Tanenhaus, M.K., & Garnsey, S.M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*(3), 285–318.

Vasishth, S., & Drenhaus, H. (2011). Locality in German. *Dialogue and Discourse, 2*(1), 59–82.

Wagers, M.W., Lau, E.F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61*(2), 206–237.

Weber-Fox, C.M., & Neville, H.J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioural evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8*(3), 231–256.

Wen, Z. (2007). Online processing of subject–verb agreement in English by first and second language speakers. In H. Caunt-Nulton, S. Kulatilake, & I.H. Woo (Eds.), *Proceedings of the 31st annual Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press, 668–679.

Wen, Z., Miyao, M., Takeda, A., Chu, W., & Schwartz, B. D. (2010). Proficiency effects and distance effects in non-native processing of English number agreement. In K. Franich, K. M. Iserman, & L. L. Keil (Eds.), *Proceedings of the 34th annual Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press, 445–456.

Williams, R.S., & Morris, R.K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology, 16*(1-2), 312–339.

Appendix I: Stimuli

Experimental sentences for the online reading experiment

1. The road(s) to the town(s) was/were closed to traffic for a full weekend, which caused frustration and increased stress levels for residents and visitors alike.

2. The writer(s) of the piece(s) was/were hired by the editor of the magazine, who was impressed by the underlying message and the innovative writing style.

3. The cheese(s) on the plate(s) was/were enjoyed only by some guests who did not have a particularly strong sense of smell. The others stayed at a safe distance.

4. The part(s) in the play(s) was/were acted brilliantly. This was admitted even by the strongest critics, who made it their business to pick on every other detail.

5. The request(s) from the graduate(s) was/were denied by the selection committee, which was known to adhere to the rules very strictly and never to make exceptions.

6. The picture(s) on the cover(s) was/were selected by the readers, who could vote for their favourites on the magazine's website a month before it was published.

7. The trail(s) along the river(s) was/were walked by nature lovers from all over the country, who were usually well equipped with proper shoes and fancy cameras.

8. The editor(s) of the newspaper(s) was/were admired by the staff, especially for the excellent time management that was needed to get the job done successfully on a daily basis.

9. The author(s) of the poem(s) was/were discovered after competing in an annual poetry contest that had uncovered many talents since it was first organised a decade ago.

10. The question(s) in the survey(s) was/were answered by only half of the respondents, which was most likely a result of the unclear phrasing or the sensitive nature of the topic.

11. The lawyer(s) on the case(s) was/were believed to be the best in the field, which explained the exorbitant hourly rate for the service.

12. The task(s) for the employee(s) was/were expected to be too demanding, which was based mainly on disappointing past performance and an apparent lack of ambition.

13. The reason(s) for the delay(s) was/were explained calmly by the airport personnel, but some of the travellers were clearly reluctant to accept the situation and shouted in response.

14. The surprise(s) for the nurse(s) was/were left on the bedside table by a favourite patient. She was happy to go home, but at the same time she was sad to say goodbye to the friendly hospital staff.

15. The skill(s) of the worker(s) was/were noticed by the manager of the company, who was pleased to find that the intensive training had been a success.

16. The member(s) of the board(s) was/were reminded that the responsibilities that came with such positions were to be taken very seriously and that any abuse of power would not be tolerated.

17. The victim(s) of the crime(s) was/were questioned by the police, who grew confident they would quickly find the offender with information this precise.

18. The painting(s) of the flower(s) was/were judged to be fake by the art expert. Admittedly, the colours seemed unnaturally bright.

19. The guard(s) of the prison(s) was/were threatened by the inmates on an almost daily basis. Fortunately, nothing serious had ever happened.

20. The border(s) of the area(s) was/were agreed upon after a long discussion between the buyer and the seller of the land, who had conflicting interests.

21. The map(s) of the island(s) was/were improved only after many tourists had complained after getting lost and missing the last boat back to the mainland.

22. The talk(s) by the scientist(s) was/were attended by many people who were not afraid to interrupt with comments or questions. Fortunately, most of them were quite interesting.

23. The sign(s) in the street(s) was/were moved by a gang of youths. This caused general confusion for motorists and cyclists, who in some cases were fined for their unlawful conduct.

24. The ad(s) on the screen(s) was/were experienced by most people as distracting rather than appealing, which was not exactly what the advertising company had hoped for.

25. The decision(s) of the judge(s) was/were reached after a long and difficult trial, during which several witnesses gave rather contradictory evidence.

26. The goal(s) of the exhibition(s) was/were achieved when the museum had welcomed its thousandth visitor and when several positive reviews had appeared in local newspapers.

27. The purpose(s) of the lecture(s) was/were expressed patiently by the professor, who had become used to students who could not fully appreciate her beloved subject.

28. The secret(s) of the employer(s) was/were discovered when the assistant overheard an indiscreet and revealing phone call while copying some documents in the next room.

29. The computer(s) in the office(s) was/were replaced with a much newer and quicker model, which was a relief to anyone who actually wanted to get some work done.

30. The expression(s) on the face(s) was/were understood by the applicant, who had known that his chances were minimal even before the interview had started.

31. The answer(s) to the message(s) was/were read out loud by the frustrated teacher, who had been disturbed by the giggling girls at the back one too many times.

32. The run(s) through the village(s) was/were cancelled three days in advance, because the weather forecast warned of fierce winds and heavy rain.

33. The girl(s) in the drawing(s) was/were captured beautifully by the young and talented artist, who usually sold her best works to save up for her own gallery.

34. The assumption(s) of the teacher(s) was/were doubted by the know-it-all students, who would rather turn to their phones and tablets for information.

35. The advantage(s) of the agreement(s) was/were acknowledged by both parties in the end, despite their long dispute over the most minor details.

36. The customer(s) of the store(s) was/were greeted by the managers, who made a slightly insincere impression with their wide smiles and friendly remarks.

37. The claim(s) of the scholar(s) was/were challenged by peers, who felt that more concrete evidence was needed before any reasonable conclusions could be drawn.

38. The gift(s) to the couple(s) was/were intended as a joke, which was not appreciated by everyone present at the engagement party. In fact, some of the guests took offence.

39. The fire(s) in the forest(s) was/were caused by extreme dryness in the summertime, which was a recurring and extremely serious problem for the developing country.

40. The appointment(s) with the adviser(s) was/were scheduled at least a month in advance, simply because many people needed financial advice in these troubled times of economic crisis.

41. The topic(s) of the chapter(s) was/were approached in a logical and comprehensible way, which was a great relief to the students studying for their exam.

42. The statement(s) of the government(s) was/were released two days after the terrible incident had taken place. Many people agreed that a response should have been broadcast much earlier.

43. The owner(s) of the building(s) was/were encouraged to finally do something about the maintenance after all tenants agreed to stop paying the rent until there would be significant improvements.

44. The confession(s) of the suspect(s) was/were claimed to be false after new evidence was obtained from surveillance cameras in the nightclub.

45. The fee(s) for the competition(s) was/were increased considerably and therefore fewer people were expected to take part.

46. The ticket(s) for the train(s) was/were left behind in the ticket machine. Unfortunately, the ticket inspector had heard that story before and did not believe it this time.

47. The election(s) for the council(s) was/were expected to attract a record turnout, since many people were dissatisfied with the current state of affairs in the city.

48. The reader(s) of the item(s) was/were encouraged to donate money to voluntary organisations active in the area affected by the devastating earthquake.

49. The solution(s) to the shortage(s) was/were proposed by a team of volunteers with years of experience. Fortunately, the problem was solved within two days.

50. The insurance(s) for the car(s) was/were paid for by the employer, who had had bad experiences with inexperienced drivers and uninsured vehicles in the past.

51. The spot(s) at the airport(s) was/were intended for people who needed a smoke to calm down before boarding their planes.

52. The cause(s) of the accident(s) was/were investigated in detail by the police, but no concrete evidence was found at the scene.

53. The letter(s) to the parent(s) was/were required to be polite and friendly. This was the school's policy even when the teenager in question had misbehaved badly.

54. The experience(s) of the user(s) was/were improved immensely when the new version of the graphics editing programme was launched.

55. The floor(s) in the hospital(s) was/were cleaned at least three times a day, which dramatically reduced the risk of infection for the surgical patients.

56. The speech(es) by the host(s) was/were delivered in such a funny way that the audience continuously roared with laughter and approval.

57. The account(s) of the battle(s) was/were written by a respected and impartial historian, who had carefully checked all facts and particulars with multiple sources.

58. The enemy(ies) of the hero(es) was/were defeated at the end of the Hollywood film, which was entirely predictable but pleasing at the same time.

59. The complaint(s) from the neighbour(s) was/were dismissed by the young man, who thought he had every right to turn up the volume for a party on a Saturday night.

60. The label(s) on the bottle(s) was/were added to warn consumers against underage drinking and the general dangers of alcohol consumption.

Comprehension questions for the experimental sentences

1. Did the people who were affected by the access problem respond calmly? (NO)

2. Did the editor of the magazine admire the writing? (YES)

3. Was the smell of all of the snacks served at the event pleasant? (NO)

4. Did the critics live up to their names by being critical of most things other than the superb acting? (YES)

5. Was the selection committee known to be very flexible? (NO)

6. Were the readers somehow involved in choosing the layout of the magazine? (YES)

7. Did the hikers often seem quite unprepared for their walks in the area? (NO)

8. Did the staff consider good time management as a very desirable leadership quality? (YES)

9. Had the poetry competition been organised just once before? (NO)

10. Could the design of the materials have been a bit better? (YES)

11. Did the service cost much less than could be expected based on the quality? (NO)

12. Was a poor past performance one of the reasons for the employer to have low expectations? (YES)

13. Did all of the travellers respond to the situation in a reasonable way? (NO)

14. Does it seem like the patient was treated well during her stay at the hospital? (YES)

15. Was the training that was offered by the company very basic? (NO)

16. Were clear instructions given about what was expected? (YES)

17. Were the police worried they would never find the offender? (NO)

18. Did the colours in the art seem quite unnatural? (YES)

19. Were the inmates always well-behaved and friendly? (NO)

20. Did it take some time before the buyer and the seller came to an agreement? (YES)

21. Does it seem likely that the tourists had a good laugh after getting lost? (NO)

22. Did the audience tend to be quite active and make contributions? (YES)

23. Did the traffic continue to function normally after the alterations were made by the youths? (NO)

24. Had the advertising company hoped for a much more positive evaluation? (YES)

25. Did the witnesses at the trial all agree on what had happened? (NO)

26. Does it seem like the reviewers liked what they saw in the museum? (YES)

27. Did all of the students seem to be interested in the subject taught by the professor? (NO)

28. Was the assistant doing something useful when he or she overheard the conversation? (YES)

29. Did the employees probably get less work done with the new model? (NO)

30. Did the applicant think beforehand that he probably wouldn't be hired? (YES)

31. Did the teacher manage to ignore the giggling girls? (NO)

32. Was the predicted weather the reason for the cancellation of the running event? (YES)

33. Did the artist tend to store her best works to hang them in her future gallery? (NO)

34. Did the students seem to prefer online sources for their information? (YES)

35. Did the parties agree immediately, without much discussion? (NO)

36. Did the behaviour of the managers seem a little fake or studied? (YES)

37. Did peers agree that the evidence was sufficiently concrete? (NO)

38. Did some of the guests seem to find the joke inappropriate? (YES)

39. Does it seem like the country usually has a heavy rainfall in the summertime? (NO)

40. Did people have to book the service quite some time in advance? (YES)

41. Did the students seem to find the materials for the exam highly complex? (NO)

42. Did many people think the authorities could have done a better job handling the situation? (YES)

43. Did the tenants seem perfectly satisfied about the maintenance? (NO)

44. Did the video recordings from the surveillance cameras reveal new information? (YES)

45. Was the number of competitors expected to be higher compared to former years? (NO)

46. Did the ticket inspector think the story was made up? (YES)

47. Were many people expected to vote because they were happy with the way things were? (NO)

48. Was the area affected by a natural disaster? (YES)

49. Did it take weeks before the problem finally got solved? (NO)

50. Did the employer have good reasons to want to be well-insured? (YES)

51. Is it true that people were not allowed to smoke anywhere before boarding? (NO)

52. Did the police conduct a thorough investigation at the scene? (YES)

53. Did the school allow strongly worded communication towards parents in case this was thought to be necessary? (NO)

54. Did the new version of the editing programme seem better than the previous ones? (YES)

55. Does it seem that a risk of infection was high because the cleaning wasn't done properly? (NO)

56. Did the audience seem to have a good time and enjoy the show? (YES)

57. Did the historian invent most of what he wrote himself? (NO)

58. Does it seem like the ending of the film was very typical and happy? (YES)

59. Did the young man probably respond in a polite and considerate way? (NO)

60. Does it seem like the producer of this beer brand is aware of the alcohol laws? (YES)

Filler sentences for the online reading experiment

61. The little café became a real tourist hotspot after the famous scene had been filmed there. Nonetheless, it had somehow retained its authenticity.

62. The wildlife photographer considered herself extremely lucky professionally. She got to see the most amazing things and call it work.

63. The powerful politician was corrupt in every sense of the word. He managed to bribe even his worst enemies.

64. The vulture may not be the prettiest bird, but it's definitely a useful creature. The African plains would be a lot messier without its cleaning service.

65. When the popular band announced it would perform at the summer festival, the tickets were sold out within the hour.

66. Vitamin D is essential for your health. Make sure to catch plenty of sunlight in the summer, but do remember to protect your skin against sunburn.

67. The talented sous-chef was promoted when he managed to impress the tough food critic with his latest taste explosion.

68. The city is known for its incredible architecture and exciting nightlife. It was named the European capital of culture in 2010.

69. The man had driven without a driving licence for 36 years before he was found out in a routine check. The story got into all the newspapers.

70. The unknown athlete was a great surprise in the Olympics. She easily beat the favourite and won three gold medals.

71. The waiter was so rude that he didn't get any tips that night. In fact, one of the customers refused to pay at all.

72. The writer's novels were real page-turners. His latest book ranked number two in the bestseller list within a week after its publication.

73. Facial hair is obviously making a comeback. All ads in men's magazines now feature models with beards and moustaches.

74. Swimming is recommended as good all-round exercise. Several muscle groups are trained and harmful effects of stress are greatly reduced.

75. They always play great obscure music on this radio station. The DJs obviously know what they're doing.

76. The stand-up comedian forgot what he was getting at midway through his story. Luckily, the man could improvise.

77. The production of the new medicine was ceased after several studies had concluded that there was no measurable effect.

78. Sleep deprivation in its chronic form can cause daytime sleepiness and clumsiness, and it can lead to weight loss or weight gain.

79. The thick fog made driving conditions dangerous. There were several small incidents, but no-one got hurt.

80. The cyclist felt a rush of pure adrenaline as he won the race. The runner-up only arrived minutes later.

81. As a result of the economic crisis, camping trips are regaining in popularity as a cheap alternative to stays in holiday homes and hotels.

82. The popularity of the series was astonishing. Even people who normally didn't like the genre couldn't wait for the new season to be aired.

83. A light to moderate offshore wind makes for ideal surf conditions, especially for beginning surfers.

84. In the spring, as the trees start to bloom and pollen becomes airborne, allergy sufferers begin their annual ritual of sniffling and sneezing.

85. The new diet recommended by the self-proclaimed lifestyle guru was challenging, but it seemed to have a lasting effect.

86. The theatre was reopened three months after the unfortunate incident where the ceiling collapsed during a performance.

87. The award-winning journalist liked taking risks to find a good story, even when that meant being in actual danger.

88. A recent large-scale study revealed that a shocking 73 percent of married people feel like they just 'settled' for their partners.

89. Social media create exciting possibilities for companies to reach the public. Commercial organisations that don't use them miss out.

90. Queuing is one of the nation's most popular pastimes. If you want to blend in, you'd better learn the rules of the queue.

91. The gorilla's escape from the enclosure attracted a lot of bad publicity for the zoo because a visitor got injured.

92. The modern art museum always displayed exciting and controversial collections that were sure to attract many visitors.

93. Keeping horses is an expensive hobby as well as a real commitment. Horses need a lot of daily care and attention.

94. The woman who owned the music shop was worried about bankruptcy. CD sales had noticeably decreased since last year.

95. In terms of flavour, there's no comparing fresh mint tea to the regular dried version. Try adding some honey and be surprised.

96. The chic restaurant had to be booked several weeks in advance, but it was definitely worth the wait. Sadly, however, the portions were not very generous.

97. Musicals are not for everyone to enjoy. Some people think it's quite unnatural when a story is conveyed in song.

98. The bakery on the square was famous for its delicious fresh pastries. With every change of season, the whole selection was replaced with yet more fabulous treats.

99. The protest against the planned new road was led by conservation groups who were concerned about a protected species of bird in the area.

100. The television interview with the highest-paid actress in Hollywood was revealing and unexpectedly disarming.

Comprehension questions for the filler sentences

61. Did the café lose a lot of customers after the scene had been filmed there?   (NO)

62. Would the photographer likely want to keep her job?   (YES)

63. Is it true that the politician was a very honest man?   (NO)

64. Do vultures have an important role on the African plains?   (YES)

65. Did it take months before the tickets were sold out?   (NO)

66. Is sunlight likely a source of vitamin D?   (YES)

67. Was the dish a disaster?   (NO)

68. Will an architecture lover probably have a good time in this city? (YES)

69. Is it true that the story didn't get any attention at all? (NO)

70. Did the unknown athlete perform well at the Olympics? (YES)

71. Were the customers pleased with the waiter's behaviour? (NO)

72. Is the writer probably quite famous? (YES)

73. Are the male models in the ads well-shaven? (NO)

74. Is swimming a good idea when you're stressed? (YES)

75. Is the music on this radio station much too mainstream? (NO)

76. Did the stand-up comedian probably solve his problem? (YES)

77. Was the medicine very effective? (NO)

78. Can sleep deprivation have an unpredictable effect on body weight? (YES)

79. Were driving conditions ideal? (NO)

80. Was the performance of the winner probably quite impressive? (YES)

81. Are camping trips becoming less and less popular? (NO)

82. Was the series a great hit? (YES)

83. Is heavy wind a great surf condition? (NO)

84. Does pollen have an unpleasant effect for allergic people? (YES)

85. Did people tend to gain a lot of weight after the diet? (NO)

86. Did the performance end in an unexpected way? (YES)

87. Was the journalist unwilling to compromise his or her safety? (NO)

88. Did the study reveal that the majority of married people haven't waited for 'true love'? (YES)

89. Is it a bad idea for companies to promote their services on social media? (NO)

90. Is queuing an important phenomenon in a specific culture? (YES)

91. Is it likely that the zoo saw an increase in visitor numbers after the incident? (NO)

92. Does it seem like the museum was popular? (YES)

93. Can you easily skip taking care of a horse for a day? (NO)

94. Does it seem like the CD industry is in trouble? (YES)

95. Is dried mint tea much tastier than the fresh version? (NO)

96. Is it likely that a customer would leave the restaurant still feeling hungry? (YES)

97. Does everyone love it when a story is sung, as in musicals?   (NO)

98. Did the bakery probably offer seasonal pastries?   (YES)

99. Was the protest led by people who didn't want a big road close to their homes? (NO)

100.Was the actress probably very open and honest in the interview?   (YES)

Items for the multiple choice test

The number of participants (out of 18) who answered a specific multiple choice question incorrectly is printed between parentheses behind each of the definitions.

1. A food made from milk, either soft or hard and often yellow or white   (0)

    a) cheese   b) a pancake   c) a dish   d) shampoo

2. A question which politely or officially asks for something   (0)

    a) an interview   *b) a request*   c) a list   d) a message

3. Something you accept as true without question or proof   (1)

    a) an invitation   b) a reason   c) a guess   *d) an assumption*

4. A statement that something is true   (0)

    a) a claim   b) a belief   c) a case   d) a reward

5. Admitting that you have done something wrong or illegal   (2)

    a) a statement   *b) a confession*   c) a joke   d) a story

6. A formal arrangement to meet or visit someone   (0)

    a) an encounter   b) a concept   *c) an appointment*   d) a date

7. A statement that something is wrong or not satisfactory   (1)

    a) a commitment   b) a discussion   c) an accusation   *d) a complaint*

8. A small piece of paper or other material that gives information about the   (4)
    thing it is fixed to

    a) a label   b) a sign   c) a sticker   d) a shape

9. Someone who has studied for and received a degree at a university   (0)

    a) a student   *b) a graduate*   c) a guard   d) a diploma

10. A piece of writing in which the words are arranged in separate lines, often   (1)
    ending in rhyme

    a) a letter   b) a success   *c) a poem*   d) a play

11. A situation in which you have to wait longer than expected   (0)

    a) an experience   b) an accident   c) a pause   *d) a delay*

12. A group of people responsible for controlling and organising a company (1)
   or organisation

   a) employers   *b) a board*   c) users   d) a government

13. An event at which for example art objects are shown to the public (3)

   a) a gallery   b) a prison   *c) an exhibition*   d) a show

14. A formal talk on a serious subject (6)

   a) a flyer   b) a topic   c) a speech   *d) a lecture*

15. A person or organization that employs people (2)

   a) an employer   b) a boss   c) a leader   d) a parent

16. A picture made with materials such as a pencil or pen (2)

   a) a painting   *b) a drawing*   c) a map   d) a device

17. A person who studies a subject in much detail, especially at a university (4)

   a) a teacher   b) a hairdresser   *c) a scholar*   d) a specialist

18. The separate parts of a book or piece of text, usually given a number or a title (0)

   a) a file   b) a cover   c) a page   *d) a chapter*

19. A person believed to have committed a crime or done something wrong (0)

   a) a suspect   b) a criminal   c) a witness   d) an author

20. A person who introduces guests and performers on television or radio (0)

   a) a producer   *b) a host*   c) an engineer   d) a journalist

21. Someone whose job is to give advice about a subject (2)

   a) a customer   b) a lawyer   *c) an adviser*   d) a coach

22. A group of people elected or chosen to make decisions and give (4)
   advice on a particular subject

   a) a club   b) a convention   c) a focus group   *d) a council*

23. A situation in which there is not enough of something (0)

   a) *a shortage*   b) absence   c) a decrease   d) an advantage

24. To respect and approve of someone or their behaviour (*10)

   a) to adore   *b) to admire*   c) to marry   d) to support

25. To decide that an organised event will not happen   (0)

    a) to imagine   b) to quit   c) to postpone   *d) to cancel*

26. To feel uncertain or unconfident about something, or to think it is not   (2)

    probable

    a) to doubt   b) to hesitate   c) to wonder   d) to compare

27. To welcome someone with words or an action   (1)

    a) to approach   *b) to greet*   c) to meet   d) to allow

28. To allow something to be shown in public or to be available for use   (5)

    a) to order   b) to confess   *c) to release*   d) to reveal

29. To arrange that an activity or event will happen at a certain time   (2)

    a) to provide   b) to decide   c) to organise   *d) to schedule*

30. To win against someone in a fight   (0)

    a) to defeat   b) to battle   c) to destruct   d) to leave

31. To decide that something or someone is not important or worth considering   (0)

    a) to refuse   *b) to dismiss*   c) to divorce   d) to praise

* Item 24 was excluded from analyses because the majority of participants (55.6%) gave an incorrect answer.

Appendix II: Statistical analyses

Several of the regression outputs referred to in chapter 5 (sections 5.2.1 and 5.2.2) can be found in this appendix.

Table 12: Regression output for the non-regression-ended first-pass reading times

Fixed effects	Estimate (log)	Std. Error	t-value	*p*-value
Intercept (condition a: group native speaker)	5.722	0.068	83.61	0.000 *
Condition b	0.140	0.055	2.54	0.002 *
Condition c	0.250	0.057	4.38	0.000 *
Condition d	0.074	0.053	1.41	0.169
Condition e	0.078	0.053	1.48	0.148
Condition f	0.041	0.049	0.84	0.141
Group learner	0.411	0.091	4.52	0.000 *
Age (centered)	-0.039	0.014	-2.83	0.008 *
Length preceding noun (centered)	0.023	0.008	2.95	0.006 *
Condition b: group learner	-0.185	0.072	-2.58	0.014 *
Condition c: group learner	-0.141	0.074	-1.90	0.066
Condition d: group learner	-0.000	0.068	0.00	0.998
Condition e: group learner	-0.016	0.067	-0.24	0.813
Condition f: group learner	0.075	0.064	1.18	0.248
Intercept (condition a: group learner)	6.133	0.061	100.33	0.000 *
Condition b	-0.045	0.048	-0.94	0.355
Condition c	0.109	0.048	2.29	0.029 *
Condition d	0.074	0.045	1.63	0.114
Condition e	0.062	0.046	1.34	0.189
Condition f	0.116	0.043	2.70	0.011 *

There were 1515 observations after the removal of regression-ended passes and data trimming.

Table 13: Regression output for the residualised non-regression-ended first-pass reading times

Fixed effects	Estimate (log)	Std. Error	t-value	*p*-value
Intercept (condition a: group native speaker)	-0.085	0.035	-2.45	0.020 *
Condition b	0.159	0.055	2.92	0.006 *
Condition c	0.223	0.055	4.05	0.000 *
Condition d	0.072	0.051	1.40	0.170
Condition e	0.070	0.052	1.34	0.190
Condition f	0.005	0.047	0.10	0.919
Group learner	0.067	0.045	1.48	0.149
Condition b: group learner	-0.183	0.072	-2.54	0.016 *
Condition c: group learner	-0.183	0.072	-2.55	0.016 *
Condition d: group learner	-0.041	0.066	-0.62	0.538
Condition e: group learner	-0.052	0.068	-0.77	0.449
Condition f: group learner	0.048	0.062	0.77	0.448
Intercept (condition a: group learner)	-0.019	0.030	-0.62	0.539
Condition b	-0.024	0.048	-0.51	0.617
Condition c	0.040	0.046	0.86	0.397
Condition d	0.030	0.044	0.67	0.498
Condition e	0.018	0.046	0.39	0.703
Condition f	0.053	0.042	1.27	0.215

There were 1515 observations after the removal of regression-ended passes and data trimming.

Table 14: Regression output for the residualised total reading times

Fixed effects	Estimate (log)	Std. Error	t-value	*p*-value
Intercept (condition a: group native speaker)	-0.053	0.042	-1.28	0.209
Condition b	0.087	0.065	1.35	0.187
Condition c	0.250	0.072	3.47	0.001 *
Condition d	0.108	0.063	1.72	0.094
Condition e	-0.046	0.069	-0.67	0.509
Condition f	-0.098	0.062	-1.57	0.126
Group learner	0.022	0.055	0.39	0.696
Condition b: group learner	-0.042	0.077	-0.55	0.586
Condition c: group learner	-0.229	0.088	-2.61	0.014 *
Condition d: group learner	-0.077	0.081	-0.95	0.349
Condition e: group learner	0.116	0.083	1.40	0.172
Condition f: group learner	0.110	0.078	1.42	0.166
Intercept (condition a: group learner)	-0.032	0.036	-0.87	0.389
Condition b	0.045	0.058	0.77	0.444
Condition c	0.021	0.064	0.32	0.748
Condition d	0.031	0.054	0.57	0.574
Condition e	0.070	0.062	1.13	0.267
Condition f	0.012	0.055	0.22	0.825

There were 1812 observations after data trimming.

Table 15: Regression output for the probability of a first-pass regression (logit)

Fixed effects	Estimate (logit)	Std. Error	z-value	p-value
Intercept (condition a: group native speaker)	-1.448	0.262	-5.53	0.000 *
Condition b	-0.603	0.442	-1.36	0.173
Condition c	-0.125	0.437	-0.29	0.775
Condition d	-0.469	0.419	-1.12	0.263
Condition e	-1.081	0.427	-2.53	0.011 *
Condition f	-1.148	0.436	-2.64	0.008 *
Group learner	-0.234	0.348	-0.67	0.501
Condition b: group learner	-0.058	0.536	-0.11	0.914
Condition c: group learner	-0.408	0.540	-0.76	0.450
Condition d: group learner	0.285	0.508	0.56	0.575
Condition e: group learner	1.282	0.497	2.58	0.001 *
Condition f: group learner	1.272	0.517	2.46	0.014 *
Intercept (condition a: group learner)	-1.693	0.239	-7.08	0.000 *
Condition b	-0.649	0.427	-1.52	0.129
Condition c	-0.527	0.426	-1.24	0.216
Condition d	-0.175	0.373	-0.47	0.640
Condition e	0.211	0.311	0.68	0.497
Condition f	0.138	0.316	0.44	0.663

Table 16: Regression output for the probability of skipping the copula first-pass (logit)

Fixed effects	Estimate (logit)	Std. Error	z-value	p-value
Intercept (condition a: group native speaker)	1.107	0.390	2.84	0.005 *
Condition b	-0.625	0.327	-1.91	0.056
Condition c	-1.333	0.334	-3.99	0.000 *
Condition d	-1.689	0.437	-3.86	0.000 *
Condition e	-1.679	0.367	-4.58	0.000 *
Condition f	-1.623	0.346	-4.69	0.000 *
Group learner	-1.965	0.521	-3.77	0.000 *
Condition b: group learner	0.656	0.429	1.53	0.126
Condition c: group learner	0.213	0.465	0.46	0.646
Condition d: group learner	0.082	0.609	0.14	0.892
Condition e: group learner	0.852	0.485	1.76	0.079
Condition f: group learner	0.085	0.490	0.17	0.862
Intercept (condition a: group learner)	-0.864	0.340	-2.54	0.011 *
Condition b	0.040	0.276	0.14	0.886
Condition c	-1.111	0.321	-3.46	0.001 *
Condition d	-1.597	0.455	-3.51	0.000 *
Condition e	-0.816	0.331	-2.47	0.014 *
Condition f	-1.534	0.363	-4.23	0.000 *