

# Universal Parsing in Second Language Sentence Comprehension<sup>1</sup>

Edson T. Miyamoto

Keywords: reading times, second language, relative clauses, subject-verb number agreement

## Abstract

This paper summarizes non-native speakers' reading-time results for two types of constructions. In the first experiment, native Chinese speakers were faster to read relative clauses in Japanese when extraction was from the subject position compared to when extraction was from the object position. In the second experiment, native Japanese speakers slowed down when reading sentences in English containing a subject-verb agreement violation. In both experiments, the results for learners are similar to what has been reported for native speakers, suggesting that non-native readers process sentences in ways that are similar to what has been reported for native readers. The paper also discusses methodological issues such as advantages of reading-time techniques over questionnaires, issues specific to non-native readers such as proficiency, as well as issues such as spillover effects, which are likely to be particularly problematic when collecting data from non-native speakers.

## 1. Learning and parsing in a second language

Traditionally, research on second language (L2) has concentrated on how people acquire knowledge of the target language. More recently, researchers have investigated L2 parsing, in other words, how learners use the knowledge they possess to process L2 sentences. Here, *parser* is used to refer to the algorithm — the way how people use linguistic knowledge to process sentences and obtain the intended meaning. For example, a learner may know that in English a subject such as *Mary* has to agree in number with a verb such as *eat* in the present tense. The question then is how learners use this knowledge in the few hundred milliseconds required to read each word in a sentence such as *Mary eats apples everyday*.

Whether language comprehension can be neatly divided into two separate components (namely, knowledge and parser) is debatable (see MacDonald, 2013, for an alternative proposal with a single component shaped by experience). But it is a useful dichotomy that helps frame the issues to be discussed here. It is also a convenient division of

---

<sup>1</sup>A shortened, revised version is to appear in Japanese as Edson T. Miyamoto & Junna Yoshida, *Dai-ni gengo-ni okeru huhentekina bun rikai mekanizumu*, in Tomohiko Shirahata & Koji Suda (Eds.), *Dai-ni gengo syuutoku monogurahu siriizu*, Vol. 2. Tokyo: Kuroshio.

labor given that traditionally different researchers have concentrated on each of those two areas. Linguists have studied language knowledge whereas psycholinguists have studied parsing.

Within this framework, learners are assumed to start with knowledge of their native language (K1) and the parser for that language (P1). Learners proceed to acquire knowledge of the target language (K2). For the most part L2 researchers have investigated how learners acquire K2, for example, how native Japanese speakers acquire knowledge of English (grammar, vocabulary, phonology and so on). A traditional question in this approach is whether learners can acquire native-like knowledge of the target language (see, for example, Schwartz & Sprouse, 1996; also Sorace, 2005, for a summary).

In this chapter, in contrast, the discussion will concentrate on the parsing mechanism. How is it that learners use their language knowledge to process sentences in the target language? The question then is about the relation between the parser for the native language (P1) and the parser for L2 (namely, P2). One possibility is that P1 and P2 are different, so learners have to learn P2 (but see Fodor, 1998, for a discussion on the logical impossibility of acquiring knowledge of language and the parsing algorithm simultaneously). Following Fodor and colleagues, we will advocate a second possibility, namely that P1 and P2 are the same. In other words, the way people use linguistic knowledge is the same regardless of the language. Therefore, it is unnecessary to learn P2, because the parsing mechanism is universal. If this view is correct, all the differences between learners and native speakers are limited to K2, the knowledge of the target language. If a learner's behavior differs from a native speaker's behavior, this is because the learner has failed to adequately acquire the relevant knowledge about the target language. As soon as the necessary knowledge is acquired, the learner should behave in a native-like manner when reading sentences in the target language.

In what follows, we will suggest that the way L2 learners process sentences is remarkably similar to what native speakers have been reported to do, thus supporting the UPH (but see Clahsen & Felser, 2006a, for a different view; also the general discussion). We will discuss two constructions, namely, relative clauses in Japanese and subject-verb agreement in English. Relative clauses are interesting because their word order varies across languages, leading to different predictions on how they strain cognitive resources such as working memory. In particular, we will suggest that native Chinese speakers learning Japanese as L2, show the same preference observed in monolingual speakers of Japanese, while preserving a different preference for Chinese.

As relative clauses allow us to investigate the effects of limited memory resources, number agreement can be used to investigate how features are manipulated. An extensive literature on number agreement in L2 English suggests that native speakers of Japanese (and other languages that do not have number agreement) do not notice agreement violations when reading L2 English. We will suggest that these previous differences between L1 and L2 readers can be explained away once two sources of variation are

factored out, namely, proficiency in the target language and so-called *spillover effects* in reading times.

## 2. The universal parsing hypothesis

The way how people process information is restricted by their cognitive resources such as working memory. For example, it is hard to memorize ten bits of information such as ten digits: 8-0-1-9-4-5-2-0-2-5. This illustrates two features of working memory, namely, information is rapidly forgotten and capacity is severely limited.

However, memorization is easier if the ten bits are grouped into three chunks: 80-1945-2025. The process is further facilitated if we create links to our long-term knowledge (e.g., a war ended in 1945) and associate the three chunks into one single complex chunk (“it will be 80 years since the end of WWII in 2025”). These operations allow us to handle complex information despite our limited working-memory capacity (see Miller, 1963).

A rough parallel can be drawn with sentence comprehension. It is hard to remember the individual words in a sentence. But we can associate words together recursively creating increasingly-complex chunks (e.g. noun phrases, verb phrases, clauses). In other words, parsing is *incremental*. We do not wait till the end of the sentence to start understanding the meaning intended. As we read each word in a sentence, we immediately associate it with the previous words read. We create links to our long-term knowledge in order to obtain a complex chunk, that is, the overall message intended. During chunking operations, information is lost and we usually only remember the overall message but not the exact words used (Anderson, 1974, and references therein).

This suggests that the human cognitive architecture determines the way how we process information. This leads to what we will refer to as the *Universal Parsing Hypothesis* (UPH), namely, that there is a single parsing algorithm for all human languages. This hypothesis rests on the assumption that the way we handle information is the same for all languages. Constraints such as those imposed by limited working-memory capacity should be the same, regardless of the language being processed. The way how we store and retrieve information should also be the same. The way we use language knowledge should also be the same even if the knowledge itself varies from language to language.

For example, incremental processing is a universal feature of language comprehension, even though in some languages the verb comes early, whereas in other languages the verb comes late in the clause. In particular, in VO (verb-object) languages such as English, the verb (e.g., its argument structure) helps predict the complements that will follow; whereas in OV languages such as Japanese, the complements (e.g., their case markers, animacy) help predict the kind of verb that will follow. In each type of language, the information used is different. Also, the category predicted is different (complements are predicted in English; verbs are predicted in Japanese). But the process is similar in that

in all languages native speakers use whatever information available to parse the fragment read (or heard) so far and predict what is to come (see Miyamoto, 2008, for a discussion on grammatical factors and incremental processing in head-final languages as opposed to head-initial languages).

### **3. Questionnaires versus reading-time experiments**

In traditional L2 research, collecting judgments about sentences is often enough for the goal of determining whether learners have acquired knowledge of some aspect of the target language. In one classic study in that tradition, English sentences were presented aurally and participants judged whether the sentences were grammatical or not (Johnson & Newport, 1989). The results indicated that L2 English learners' ability to spot ungrammatical sentences depended on their age of arrival in the United States. Learners' accuracy showed a gentle decline in lockstep with age of arrival up until around puberty. Later arrivals had much steeper declines on average.

One way of summarizing these findings is that average accuracy was high for learners whose arrival preceded puberty, and average accuracy was low for learners whose arrival occurred after puberty. But this is one of those situations in which averages are not enough to clarify the nature of the results. Another way of characterizing the results is to say that the standard deviation, in other words, the between-subject variation, was small for early arrivals, and large for late arrivals. This latter side of the results provide compelling support for a critical period in language acquisition, in other words, that biological factors determine a cutoff age until which learning a language occurs naturally (see Hyltenstam & Abrahamsson, 2005, for a summary).

When presented with this type of result, L2 learners may feel despondent as they may convince themselves that they are too old to learn a new language. Just presenting the averages before and after puberty may indeed reinforce this feeling of hopelessness. But subject variation tells a different story. Before puberty, language learning is effortless, and unimpaired children are assured to acquire the language they are exposed to. After puberty, adults can learn a new language but explicit instruction is often required. Eventual proficiency will vary considerably, as factors other than age weigh in. In other words, it is not all hopeless doom for the late learners, and there is always room for hard work to counter the loss of neural plasticity.

It should be clear from this short discussion that questionnaire data and so-called off-line judgments (i.e., judgments after the whole sentence is read or heard) provide useful information about the range and limitations in the knowledge that learners possess. However, for present purposes, questionnaires have one crucial limitation in that they only measure the knowledge that the learner possesses, but they cannot determine how well the learner can use such a knowledge.

In contrast, reading times allow us to gauge how well learners use their knowledge of L2. In the following, we summarize reading-time results for L2 learners that are





memory, therefore neutralizing the effects of dependency-length and working-memory load, thus allowing for other factors to determine processing difficulty (Miyamoto, 2016).

Like in Japanese and Korean, RCs are prenominal in Chinese, therefore closure should occur at the end of the embedded clause in this language as well. However, there are two confounds in Chinese. First, word order is mixed and many constructions are head-initial, therefore closure cannot always be performed at the phrasal head. Second, even though RCs are head-final, the embedded-clause end is marked with a functional word which may be read too quickly for closure to occur consistently. These differences may explain the discrepant results reported for Chinese RCs (see Vasishth et al., 2013, and references therein).

In Japanese, closure can be consistently performed at the phrasal head since all constructions are head final. In RCs in particular, the embedded-clause end is marked with an inflected predicate, which usually takes long enough to read in order for closure to take place. Moreover, closure is assumed to be a universal operation of the parsing mechanism. Therefore, advanced L2 learners of Japanese should perform closure at the end of the embedded clause, and consequently making length dependency irrelevant when the modified noun is read. In this case, like native Japanese speakers, learners should favor SRCs when reading in Japanese. A self-paced reading experiment was conducted to test this prediction (Tsujino & Miyamoto, 2016).

#### 4.1 Experiment 1: RCs in L2 Japanese

A group of 15 native Chinese speakers participated in the experiment. They were all advanced learners who had attended the most advanced classes of Japanese offered at the University of Tsukuba.

Participants read dialogues between two people, X and Y, who were chatting while looking at photographs.

- (5) a. X-san: Kono hito-wa daredesuka?  
X: this person-top who-is  
“Who is this person?”
- b. (SRC) Y-san: Sato-kun-o aisiteita onnanoko desu.  
(ORC) Y-san: Sato-kun-ga aisiteita onnanoko desu.  
Y: Sato-acc/-nom loved girl is  
SRC: “It is the girl who loved Sato.”  
ORC: “It is the girl who Sato loved.”

The crucial sentence is Y’s response in (5b), which contained an SRC or an ORC. The dialogue context was used to avoid ambiguity and prevent interpretations other than RCs while participants read Y’s reply (see Miyamoto & Tsujino, 2016, for details).

A total of 24 pairs of dialogues were created but results for 15 pairs are reported (similar reading-time trends were observed when all items were included). These 15 items were chosen based on the results of two norming studies conducted with native Japanese speakers. First, according to a plausibility questionnaire, the meaning intended were similarly natural (SRC: 1.54; ORC: 1.65; cumulative link model:  $p = .384$ ). Second, in a fragment-completion questionnaire, both SRCs and ORCS elicited more than 98% of appropriate completions.

The 24 test items were distributed into two Latin Square lists so that each list contained only one version of each pair of items. Each participant saw one list shown together with the 24 filler items.

The twenty-four filler dialogues were created following the general format of the test trials. The RC was replaced by constructions such as genitives (e.g., *Yamada-san-no kouhai-no Yuuta-kun desu* “it is Yamada’s junior, Yuuta.”) and multiple sentences (*Kasyu-no Ooki-san desu. Endo-san-no siriiai-desu.* “It is the singer, Ooki. It is Endo’s acquaintance.”).

A non-cumulative sentence-by-sentence self-paced reading experiment was conducted in Japanese (using Doug Rohde’s Linger program). A sentence-by-sentence presentation was adopted because L2 readers may use situation-specific strategies to read sentences and are only able to build representations when allowed to backtrack repeatedly relying on processes dissimilar to those used by L1 readers. If correct, this should predict that reading times to whole sentences would not reveal the SRC preference observed with native Japanese readers.

Reading times were skewed, therefore they were  $(-1/\sqrt[3]{RT})$ -transformed following Box-Cox analyses (package MASS on R, Venables & Ripley, 2002; similar trends were observed with the raw data as well as with a simple inverse transformation). Backward selection was used to choose the linear mixed-effects models reported. All statistical analyses reported in this chapter were conducted on R (R Core Team, 2017).

#### 4.1.1 Results and discussion

Reading times were analyzed using linear mixed-effects models. There were no spurious RC-related differences in the reading times to the initial question in (5a), “who is this person?” ( $p = .4$ ). The crucial sentence in (5b) was read faster when it included an SRC rather than an ORC ( $p = .013$ ) replicating similar trends for L1 Japanese (Miyamoto & Tsujino, 2016).

The SRC advantage held even though three other factors also contributed reliably in the model. The first effect was that the higher the plausibility ratings (using item means obtained from the norming questionnaire), the faster the reading times ( $p = .049$ ). The second effect was that the longer the sentence (measured in number of characters), the slower the reading times ( $p = .04$ ). These two effects are unsurprising and jibe well with the intuition that short plausible sentences should be read faster.

The third effect was that participants got faster along the experiment as they saw more trials ( $p < .003$ ). This is also unsurprising in that it suggests that participants got faster as they got used to the experimental procedure.

The results indicate that the L1 Chinese speakers read SRCs faster than ORCs, similar to what has been reported for L1 Japanese speakers. One possibility is that these Chinese speakers are exceptional and prefer SRCs in all languages, including Chinese. Therefore, an experiment was conducted in Chinese to counter this possibility. After participating in the Japanese experiment, the native Chinese speakers participated in a self-paced reading experiment in Chinese, where they were faster to read ORCs than SRCs, replicating the original study in Chinese (Gibson & Wu, 2013; see Tsujino & Miyamoto, 2016, for details).

In sum, the native Chinese speakers were faster to read ORCs in Chinese, but faster to read SRCs in Japanese. This suggests that L2 learners can process complex syntactic structures such as RCs in their L2 (i.e., Japanese) while maintaining their L1 parsing biases similar to monolingual readers' preferences (in this case Chinese monolinguals). Under the assumption that there is a single parser for all languages, this conclusion is unsurprising. The parsing algorithm for Chinese and Japanese being the same, Chinese speakers reading L2 Japanese, should show parsing preferences similar to those of native speakers as long as their knowledge of Japanese is advanced enough.

The result is preliminary in two respects. First, it does not take learners' detailed proficiency scores as a factor when analyzing the results. Second, only reading times to whole sentences were measured; but it would be of interest to determine whether learners slow down at the modified noun or immediately thereafter, as has been reported for native Japanese speakers. Experiments under preparation address those issues. In the following sections, we discuss a different type of construction for which reading times were measured for individual words and analyzed by including learners' proficiency scores.

## 5. Word-by-word reading times

In the previous section, we reported reading times to whole sentences and claimed that learners' overall preferences were similar to what has been reported for native Japanese speakers. From here on, we will be interested in what happens at a more fine-grained level, as each word in the sentence is read in L2 English. The assumption is that readers take longer when they read words that are somehow difficult to process. For example, we can collect reading times to ungrammatical sentences such as (6a) and to grammatical sentences such as (6b).

- (6) a. \* The cake were baked for forty minutes. (ungrammatical)  
b. The cakes were baked for forty minutes. (grammatical)

We predict that the reading times to the first word in (6a) and in (6b) should be roughly the same because the same word, *the*, is read in both sentences. At the second word, *cakes* in (6b) may be read more slowly than *cake* in (6a) (see Wagers, Lau & Phillips, 2009, for such a result for native English speakers). There are various possible explanations for such a slowdown. For example, compared to the singular form, plural nouns entail more complex contexts, are usually less frequent, and longer (in this example, the plural form has an extra letter, the plural marker *s*).

The crucial comparison, that is, the comparison to determine readers' sensitivity to ungrammatical agreements is at the third word. The question of interest here is whether reading times to *were* in the ungrammatical (6a) are longer than the reading times to the same word in the grammatical (6b). This would provide evidence that readers are keeping track of agreement relations and they slow down when there is a violation.

Note that it is not enough to only have sentences like (6a). In order to determine whether reading this sentence is slow, we need to compare its reading times to those of some control sentence. In the example above, (6b) provides a straightforward baseline.

It is also worth noting that, instead of comparing (6ab), we could compare a different pair of sentences such as the following.

- (7) a. \* The cakes was baked for forty minutes. (ungrammatical)  
b. The cake was baked for forty minutes. (grammatical)

The prediction is again that the ungrammatical (7a) should be slower than the grammatical (7b) when the verb is read. The problem with this comparison is that it has a confound, namely, plural nouns such as *cakes* tend to be read more slowly than their singular forms and this slowdown can persist in the words following the noun. Therefore, even if we detect a slowdown at *was* in (7a) we would not be able to tell with certainty whether this slowdown is due to the agreement violation or whether it is due to the plural noun that preceded it (but see Wagers, Lau & Phillips, 2009, for analyses using covariates in mixed-effects models, that can address such problems).

This discussion illustrates two points. First, even when we observe the difference that we predicted, we still need to make sure that there are no alternative ways of explaining the slowdown. This can be particularly tricky when earlier differences can contaminate the reading times of later regions. Therefore, we should choose the sentences we are comparing carefully and take into consideration what may happen as each word in the sentence is read.

The second point is that we can make very specific predictions. We can not only predict that sentence (6a) should be read more slowly than sentence (6b), but we can also predict the point in the sentence where the slowdown should occur. In (6a), we predict that the slowdown should occur at the third word since this is the first point where agreement can be checked. This prediction is not always exact, it is possible that the slowdown is

observed at the target word (the verb in the examples above) as predicted, but it may persist for a couple words thereafter.

It is also possible that the slowdown is not observed at the target word, and is only observed at the following word (e.g. *baked*) and later. There are various reasons why this kind of delayed effect may be observed. In the examples above, the target words are the verbs *was* and *were*. These words are very frequent, and tend to be read quickly, making it difficult to detect slowdowns when they are read. We can nevertheless predict that the slowdown of interest (i.e., the slowdown related to agreement) should start at the verb or immediately thereafter.

In the following, we will use such reading-time comparisons to discuss whether learners slow down when reading sentences in ways that are comparable to what has been reported for native speakers.

## 6. Number agreement

Number agreement in English is interesting because it is relatively simple and advance learners are likely to know it well. Moreover, number agreement is often redundant as it is unnecessary to understand the meaning intended; therefore, learners may ignore it assuming it will have little detrimental effect on comprehension. Although the knowledge involved is simple, agreement allows us to check whether L2 sentence comprehension proceeds correctly in steps such as the following.

- (8) a. Two items (the head noun and the verb) have to be stored in working memory.
- b. Each of the two items has to be represented appropriately (e.g., with correct number feature).
- c. The syntactic representation has to be detailed enough so that it is clear which two items are relevant.
- d. The two relevant items can be correctly retrieved so that their number features can be checked against each other to make sure that they match.

Number agreement in L1 English has been extensively studied. For example, native English speakers are slower to read sentences similar to (9a) compared to sentences such as (9b) (see Wagers, Lau & Phillips, 2009, and references therein for related results and discussion on the comprehension and production of such constructions, especially in relation to retrieval of the relevant noun in step (8d) in so-called *attraction* phenomena).

- (9) a. \* The cake with the cream were baked for forty minutes. (ungrammatical)
- b. The cakes with the cream were baked for forty minutes. (grammatical)

The UPH predicts that L2 learners of English should display a similar slowdown as long as they have acquired knowledge of subject-verb agreement. To avoid the possibility that learners are transferring from their native language, we will concentrate on learners whose native languages do not have number agreement (e.g. Chinese and Japanese). Moreover, we will restrict the discussion to late learners, in other words, those first exposed to the L2 at around puberty or later (see Hyltenstam & Abrahamsson, 2005, for a summary of differences depending on age that learning begins).

There are various possible reasons for learners to fail to process agreement relations correctly. In one scenario, the learner has not acquired the necessary knowledge. If this is the case, this would still be compatible with the UPH and we just need to make sure that the learners tested are advanced enough and know the grammatical phenomenon investigated. There have been claims that L2 learners are unable to learn functional morphology that is not available in their L1 (Jiang, Novokshanova, Masuda, & Wang, 2011; *inter alia*). But this claim has been falsified by results using agreement inside noun phrases (Wen, Miyao, Takeda, Chu, & Schwartz, 2010, who also suggest that past studies may have failed to detect differences in simple constructions because proficiency scores were not up to date). Therefore, we assume that L2 learners are able to represent number features appropriately in items in working memory as described in steps (8ab).

But it is conceivable that the way learners process L2 sentences is inherently different from what native speakers do. No matter how much learners study or are exposed to the L2, they are unable to process sentences in a native-like manner. In particular, learners have been claimed to build representations that are *shallow*, in other words they are simplified, imperfect and just good enough to perform the task at hand (see Clahsen & Felser, 2006a, for a detailed discussion). In other words, L2 readers would fail in step (8c).

How simple such shallow representations are in general is unclear, but previous studies on L2 agreement in English have argued that learners are unable to keep track of subject-verb relations (in this case, between the head noun *cake* and the verb), when there are complex intervening structures such as the prepositional phrase (PP) *with the cream* as in (9) (Wen et al., 2010; *inter alia*). In particular, it is conceivable that learners build a flat structure where the hierarchical relation between the head noun *cake* and the PP *with the cream* is not built. In this case, learners may try to associate the verb with the closest noun, *cream*, and as a consequence the two sentences should be equally slow (because in neither case the verb form is compatible with *cream*) or equally fast (because the correct verb form is not retrieved and is not checked against the number feature of the noun).

There are a number of reasons why learners may rely on shallow representations. One possibility is that learners are always inefficient managing cognitive resources such as working memory when processing L2 sentences, thus forcing learners to rely on simplified representations. Another possibility is that learners have difficulty accessing L2 knowledge. In particular, learners have difficulty retrieving the relevant information (e.g., that *was* is the correct singular form for the verb in this case). When given enough

time, learners may be able to say what the correct form of the verb should be. But they are unable to retrieve this type of information rapidly enough to be able to use it when they are reading a sentence. (Learners can read a couple of words per second, and this might be much too fast for them to be able to retrieve the correct verb form.) In this case, learners may just ignore the verb form and concentrate on the meaning intended. This is particularly true for agreement, which is often redundant, unnecessary to understand the meaning of sentences.

## 6.1 Experiment 2: number agreement in L2 English

An experiment was conducted to test whether learners slow down then they read the verb in an ungrammatical sentence such as (9a) (see Wilson & Miyamoto, 2015, for details). Such a slowdown would suggest that learners are sensitive to ungrammatical agreement even in circumstances where agreement is unnecessary to understand the meaning of the sentence.

Twenty-six native Japanese speakers, undergraduates at the University of Tsukuba, who started learning English when they were 11 years or older participated in the experiment.

Sixteen pairs of sentences such as those in (10) (repeated from (9)) were prepared based on stimuli used in past L1 experiments (Pearlmutter, Garnsey & Bock, 1999; Wagers, Lau & Phillips, 2009; for discussions on related L2 results, see Yamada & Hirose, 2012; and Jiang, 2004, Experiment 2 in particular, which included the same construction investigated here but failed to detect differences related to the agreement violation).

- (10) Regions:            1   2   3   4   5   6   7   8   9   10  
a. Ungrammatical: The cake with the cream were baked for forty minutes.  
b. Grammatical:    The cakes with the cream were baked for forty minutes.

The vocabulary used, especially in the crucial nouns (e.g., *cake* and *cream*) were chosen based on textbooks used in classes that students attended and on the judgment of people with knowledge of the participants' vocabulary level.

The sentence pairs were the same except for the first noun, which was singular in the ungrammatical condition. The 16 pairs of sentences were distributed according to a Latin Square design such that each participant saw eight grammatical and eight ungrammatical sentences, interspersed with 80 filler sentences. All filler sentences were grammatical, so as to minimize the possibility that participants would notice that ungrammaticality was being tested.

After the reading-time experiment, participants answered a C-test, which is a fill-in-the-blanks test, where the second half of every other word is missing. In the C-test used (from Babaii & Shahri, 2010), the maximum score was 100 points (each word completed correctly scored as one point). The C-test scores were used as a measure of proficiency.

### 6.1.1 Procedure and analysis

A moving window non-cumulative self-paced reading experiment was conducted using Dough Rohde's Linger software (<http://tedlab.mit.edu/dr/Linger>). Button presses in lieu of reading times were recorded for each word. Each sentence was followed by a true/false comprehension question, which was never related to number agreement, so as not to draw participants' attention to the aim of the experiment. Reading time data of the trial were discarded if the participant's answer to the comprehension question was incorrect.

Reading times were analyzed using mixed-effects models, constructed using backward selection (see Wilson & Miyamoto, 2015, for further details). The independent variables were grammaticality, C-test score, and word length wherever it differed across the two conditions.

### 6.1.2 Results

Results for the comprehension question were as follows. There was no difference between the grammatical condition (93.3%) and the ungrammatical condition (92.3%) (logit mixed-effects model:  $p_s > .2$ ; individual ranges: 75%–100% for the 16 test items, 77%–97.9% for all 96 items including the 80 filler items).

Results for reading times per region were as follows. In the second region, singular nouns (e.g. *cake*) were read faster than plural nouns (*cakes*;  $\beta = -390$ ,  $p < .001$ ; replicating Jiang, 2004; also, Wagers, Lau & Phillips, 2009, for similar trends for native speakers). In this region, there was also a main effect of C-test, as participants with higher C-test scores read faster ( $\beta = -10.1$ ,  $p = .007$ ; i.e., reading time was 10.1 ms faster for each point increase in the C-test score; a similar trend was observed in region 7, *baked*:  $\beta = -9.2$ ,  $p = .002$ ).

The critical region was the verb *were*. However, no reliable differences were observed at this point (see Wagers, Lau & Phillips, 2009, for a comparable spillover pattern with native speakers). In the following region (*baked*), there was a significant interaction between grammaticality and C-test ( $\beta = 6.6$ ,  $p = .018$ ). This interaction indicates that the higher the C-test score, the larger the slowdown in the ungrammatical sentence (more specifically, for each point increase in the C-test score, the slowdown in the ungrammatical condition (9a) in relation to the grammatical condition (9b) was 6.6 ms larger).

An ancillary analysis was conducted to facilitate interpretation of this interaction, the continuous variable C-test was split into two levels with 13 participants each around the median score, and pairwise analyses were conducted. When C-test was low, there was no significant effect of grammaticality ( $\beta = -83.02$ ,  $p > .1$ ). When the C-test was high, ungrammatical sentences were read more slowly than grammatical sentences ( $\beta = 173.19$ ,  $p = .007$ ).

A previous study did not detect any differences at the verb when using a similar

construction (Jiang, 2004, Experiment 2), but this could be because it did not take proficiency into account when analyzing the data, thus increasing in-between subject variation. Indeed, when proficiency scores were not included, agreement violation did not lead to reliable differences at *baked* in the present experiment either.

## 6.2 Discussion

Proficient L2 readers were found to be sensitive to ungrammatical agreement even though their native language, Japanese, does not mark number agreement between subject and verb (see Yamada & Hirose, 2012, for related discussion using related constructions). The participants were not required to make grammaticality judgments and, in any case, the vast majority of the sentences used were grammatical (each participant saw only 8 ungrammatical test sentences, while the remaining 88 sentences were all grammatical), therefore it is unlikely that they were purposefully looking for ungrammaticalities.

In the following sections, we highlight three aspects of the results. First, the results indicate that L2 readers are able to produce representations that are more sophisticated than predicted by shallow parsing (Clahsen & Felser, 2006a). Second, proficiency is a crucial factor in reading-time analyses. Third, spillovers are an important confound that needs to be taken into consideration, especially in L2 readers' data.

### 6.2.1 Good enough is not good enough for L2 readers

If participants were only building simple representations just good enough to get past the task at hand as has been claimed in the past (Clahsen & Felser, 2006a), agreement relations could have been ignored, since they were unnecessary to answer the simple comprehension question asked after each sentence in the experiment. In fact, subject-verb agreement was unnecessary to understand the meaning of any of the 96 sentences used. Nevertheless, the L2 readers were slower to read the ungrammatical agreement sentences, thus suggesting that they were in fact keeping track of plural markers and agreement relations.

However, there is an alternative interpretation that could potentially explain the results. It is possible that L2 readers keep track of the number marker in the previous nouns and as long as one of them is marked plural, the verb must be plural as well. This predicts that a fragment such as *the cake with the creams was* should lead to slow reading times. However, similar to what has been reported for native speakers (Wagers, Lau & Phillips, 2009), results from ongoing experiments suggest that apart from slow reading times to the plural noun *creams*, there is no slow reading times at the verb.

Therefore, participants must have built syntactic structures detailed enough to be able to determine that in the fragment *the cake with the cream was*, the relevant noun for agreement with the verb is *cake* and not *cream*. Contrary to previous claims (Wen et al., 2010, and references therein), learners were able to check agreement between verb and subject, even though a fairly complex PP (*with the cream*) intervened. In general, all

the steps in (8) must have been successful. To what extent these steps are really similar to native speakers' processing of agreement remains open to the extent that we did not provide detailed evidence on the retrieval of the relevant noun in (8d). But preliminary results of on-going experiments suggest that even attraction phenomena in L2 are similar to what has been reported in previous L1 literature when spillover effects are removed (Wagers, Lau & Phillips, 2009).

### 6.2.2 Proficiency

The crucial slowdown was only detected for the proficient readers, those that scored high in the C-test. The slowdown was not detected for the readers who had low scores in the C-test. This characterization is a simplification in that it divides the participants into two categories (high versus low). This ancillary analysis was reported to facilitate comparison with previous reports, which used such partitions. But the mixed-model analyses suggest that this dichotomy is likely to be a simplification as the reading-time difference increased as the C-test score increased. In other words, sensitivity to ungrammatical agreement (i.e., slow reading times to ungrammatical agreement) increased with proficiency.

Whether we treat reading times as a function of proficiency measured as a categorical high/low variable or as a continuous variable, the present experiment suggests that it is likely that previous null results (Jiang, 2004; Jiang et al., 2011; *inter alia*) did not detect crucial differences because the participants' proficiency was not high enough or because their proficiency scores were not included in the data analyses (see Wen et al., 2010, for related discussion). In other words, there is no conflict between our results and previous literature.

If correct, this conclusion suggests that proficiency is a crucial factor when analyzing reading-time data. The group of participants tested was fairly homogeneous. Even when we restricted analyses to first-year undergraduates registered in the same major at the University of Tsukuba, we still found that considerable variation in their reading times were explained by their C-test scores (Wilson & Miyamoto, 2015).

This raises the question as to how to measure proficiency. The C-test had a high correlation with reading times, but not as high as more well-established measurements such as the TOEFL-ITP<sup>2</sup> (see Wilson & Miyamoto, 2015, for some reading-time variation that may be explained by the C-test scores but not by the TOEFL-ITP). The great advantage of the C-test is that it can be conducted in a short time (in the present case, participants were given 15 minutes to answer, based on pre-tests), whereas the whole TOEFL-ITP can take almost two hours and even the most relevant section on reading comprehension can take close to an hour. This is particularly important given previous claims that proficiency should be measured in the same section as the reading times

---

<sup>2</sup>See [https://www.ets.org/toefl\\_itp](https://www.ets.org/toefl_itp)

because L2 proficiency can fluctuate with time (Wen et al., 2010).

### 6.2.3 Spillovers

As we discussed earlier, before running a reading-time experiment, we can predict where in the sentence a slowdown is likely to occur. However, it is not uncommon for difficulty at a word to persist and for slow reading times to be observed at the following word or words. In the present experiment, the difficulty was predicted for region 6 (the verb *were*), but it was only observed at region 7 (*baked*; see Wagers, Lau & Phillips, 2009, for similar trends for native English speakers). One possibility is that the verb *to be* is read quickly making it difficult to detect differences at this point, in other words, it is a *floor effect*. This implies that the readers may have advanced to the next word before completely finishing processing the verb *to be*, thus difficulty associated with this verb was only reflected at the following word, *baked*.

This kind of *spillover effect*, where difficulty at one point is observed at later points, requires us to be careful when choosing the sentences to be compared (see the earlier discussion on the sentences in (7)). We chose the sentences in (10) to avoid a possible confound from such spillovers as in the following examples.

- (11) a. Ungrammatical: The cakes with the cream was baked for forty minutes.  
b. Grammatical: The cake with the cream was baked for forty minutes.

The prediction is that *was* should be slower in the ungrammatical sentence in (11a). However, the plural noun *cakes* is likely to be read more slowly and this slowdown may persist. Therefore, even if we detected slow reading times to the verb *was* in (11a), we would not be sure whether it was due to the ungrammatical agreement, to the plural noun, or perhaps both.

There is also the possibility that we would not find a difference between the two sentences because readers incorrectly associated *was* with the intervening noun *cream* (but ongoing experiments suggest that this is not the case with L2 readers; see Wagers, Lau & Phillips, 2009, for data on native English speakers).

In general, spillovers can occur in reading-time experiments with native as well as L2 readers. But L2 readers may be slower to recover from processing a difficult word, therefore spillovers may be more common than with native readers. Another possibility is that the same L2 reader may not be as consistent as native readers, perhaps because of lack of practice reading L2 texts. Therefore, L2 readers may display more random fluctuations than native readers, and their data may require extra care when analyzed. For example, assume that an L2 reader started reading a sentence like (11a) more slowly than usual. Such a slowdown is unlikely to be related to the nature of the first word, the article *the*, and is due to the nature of L2 reading having more random fluctuations. Such a slowdown could persist for a few words and we would like to eliminate it since it is not of interest when investigating ungrammatical agreement. One way of addressing this problem is

to include the reading times to the first word as a factor in the analyses of the following words (see Wagers, Lau & Phillips, 2009, for such an analysis with native speakers' data). In our experience, this kind of analysis can clean up L2 reading data and make it easier to detect the difference of interest (e.g., agreement violations).

Admitting that L2 reading can be noisier is not the same as saying that it is inherently different from L1 reading. As mentioned earlier, L2 readers have less experience reading L2 texts, therefore it may be more laborious to access L2 knowledge, and L2 reading may be more tiring, thus leading to random fluctuations. Perhaps more crucial is that proficiency varies more among L2 learners than among native speakers (see earlier discussion on Johnson & Newport, 1989), therefore more variation in reading-time patterns is to be expected and it should be addressed before comparing the results with native speakers'.

## **7. General Discussion**

This paper summarized results for two experiments suggesting that L2 readers slow down in ways that are similar to what has been reported for L1 readers. It did not matter whether we looked at sentence-level reading times (Experiment 1) or at word-level reading times (Experiment 2), in both cases L2 reading times were as predicted based on what we know from the L1 literature.

The results summarized here are at odds with previous studies that did not find reliable differences (see Clahsen & Felser, 2006a, for an extensive summary of null results in the L2 literature). But we do not believe that there is a contradiction. It is worth emphasizing that the vast majority of past claims were based on null results. These studies reported a failure to find some reliable difference in L2 participants' reading patterns, whereas L1 participants did show sensitivity. This contrast was then used to argue that L2 readers are unable to build detailed representations.

But it is possible that these past null results were due to some interfering factor that was absent or less prevalent in the L1 readers. We suggested a few possibilities. First, proficiency is likely to be important and analyses should include scores as a factor in the reading-time analyses, especially when looking at word-by-word results. Second, L2 reading is likely to be noisier. For example, L2 readers are less experienced and may take longer to recover from difficulty, therefore spillover is likely to be more widespread than in L1 reading.

Although some past results took proficiency in consideration (for example, by only considering high-proficiency readers, as in Clahsen & Felser, 2006a), the evaluations may have been too coarse (or included listening scores which are likely to mask relevant differences; see Wilson & Miyamoto, 2015, for the low explanatory power of listening scores when analyzing reading times). Moreover, most past analyses did not include proficiency scores as factors in the reading-time analyses. Neither were reading-time covariates included to remove spillovers from previous regions. The inclusion of

such factors would be especially relevant when testing for subtle differences such as intermediate-trace effects (as in Marinis, Roberts, Felser & Clahsen, 2005).

Note that claiming that L1 and L2 reading are similar does not entail that they are necessarily identical. Brain areas may differ more for late than early learners (e.g., Kim, Relkin, Lee & Hirsch, 1997). But such differences may be related to how knowledge is stored rather than to the parsing algorithm. In other words, the parsing algorithm may be preserved even if it has to tap different brain structures to retrieve the necessary knowledge.

## **7.1 Shallow structures revisited**

According to the shallow structure hypothesis (SSH; Clahsen & Felser, 2006abc), L2 readers are only able to build simplified representations based on adjacent words and lexico-semantics. Therefore, constructions as the ones discussed here are a good test since they rely on non-adjacent relations.

In general, long-distance dependencies (those between constituents in different clauses) should be beyond the capacity of L2 readers (Clahsen & Felser, 2006c). RCs in particular are an often-discussed example of such long-distance dependencies given the relation between the modified noun and the extraction position inside the embedded clause. If L2 readers are unable to build inter-clausal dependencies, they have to rely on factors such as lexical information and world knowledge to determine the meaning intended. The reading-time results in Experiment 1 were restricted to items that were neutral according to the results of two norming questionnaires, therefore factors such as plausibility biases and expectation of how the fragment will proceed are unlikely to have helped participants process the RCs. Consequently, these factors cannot explain why native Chinese speakers favored SRCs when reading L2 Japanese.

It is likely that the results of Experiment 2 are also beyond what the SSH can explain given the PP intervening between the noun and the verb, which required keeping track of non-adjacent syntactic relations. Clearly, it cannot be reduced to lexico-semantic information. Moreover, the agreement relation was redundant and did not help in the comprehension of the sentences.

Although the SSH seems vastly different from what we have discussed here, there are in fact points in common. The SSH seems to assume that the L2 deficit is not in parsing, but rather in the nature of L2 knowledge. Although acknowledging the possibility of different parsers, the option Clahsen and Felser favor is of a single universal parser so that “the same parsing mechanisms that are used in L1 processing (such as minimal attachment, recency, or the active filler strategy) are also available in L2 processing, but that their application is restricted due to the knowledge source that feeds the structural parser, the L2 grammar, being incomplete, divergent, or of a form that makes it unsuitable for parsing” (Clahsen & Felser, 2006b, p. 117). That is, we agree with those authors in that any differences in L2 reading patterns must stem from L2 knowledge. It also seems

reasonable to assume that because of the imperfect nature of their knowledge, L2 readers rely on shallow structures more often than L1 readers usually do. But the SSH seems to assume that L2 knowledge can never approach the nature of L1 knowledge (see Schwartz & Sprouse, 1996, for the view that L1-like levels are attainable even if not guaranteed for L2 learners), whereas we assume that as their knowledge improves, their behavior should approach that of native speakers’.

We believe the results reported in this chapter support our point of view. In Experiment 2 in particular, we observed that sensitivity to ungrammatical agreement increased with proficiency as measured by a C-test questionnaire. More specifically, for each point increase in the C-test, the ungrammatical condition was predicted to be read more slowly by an extra 6.6 ms. The results suggest that as proficiency increases, L2 reading patterns converge to what we observe in L1 readers.

In sum, the SSH is unlikely to explain the results of Experiments 1 and 2 reported here. Moreover, the SSH is also unlikely to explain past results that were supposed to support it. Extraction effects in long-distance dependencies (as reported in Marinis et al., 2005; see also Clahsen & Felser, 2006ac) have been explained by assuming that L2 readers associate a *wh*-phrase directly to a subcategorizing verb in a deeply-embedded clause based on lexico-semantics. However, the hallmark of *wh*-phrases is that they require an extraction position to be interpreted. For native speakers, the assumption is that this requirement is stored in working memory when the *wh*-phrase is read and it is satisfied as soon as a grammatical position is found later in the sentence (see the *active filler strategy* in Frazier & Clifton, 1989). This requirement for an extraction position is clearly syntactic in nature and its satisfaction also involves syntactic constraints. It is unclear how lexico-semantics could account for the processing of such a dependency across multiple clauses as discussed in the next section.

## 7.2 Only detailed representations are good enough (for proficient L2 readers)

This section provides a detailed discussion of the following examples (from Marinis et al., 2005) to show that the SSH is indeed insufficient to explain the comprehension of long-distance dependencies.

- (12) a. The nurse who the doctor argued that the rude patient had angered is refusing to work late.
- b. The nurse who the doctor’s argument about the rude patient had angered is refusing to work late.

In both sentences, the *wh*-phrase *who* has to be associated with an extraction position, that is, the object position of the predicate *had angered*. But, under some grammar formalisms (Chomsky, 1995, and references therein), this dependency is broken into

two dependencies with an intermediate landing position at *that* in (12a), but not at *about* in (12b). In other words, there is one long *wh*-dependency in (12b) (from *who* to the extraction position) and two short dependencies in (12a) (from *who* to *that*, and from *that* to the extraction position). Native English speakers' reading latencies were compatible with this distinction replicating the original results (Gibson & Warren, 2004) as summarized next.

- (13) a. The word *that* was read more slowly in (12a) than *about* in (12b), compatible with the assumption that a landing site was processed in (12a).
- b. At *had angered*, both (12ab) took longer than comparable control sentences without extractions (the controls are not shown here), suggesting that there was a cost creating the *wh*-dependency at this point.
- c. The cost of the dependency created at *had angered* was not equally large for (12ab) as the predicate was read faster in (12a) than in (12b) suggesting that the dependency was longer in the second sentence. No such a difference was found in the two control conditions without extraction, thus leading to a statistically reliable interaction at *had angered*.

Native speakers of Chinese, Japanese, German and Greek failed to display any difference at *that* in comparison to *about* (contrary to native speakers' result in (13a)). There was also no reliable interaction as summarized in (13c). The only reliable difference was the overall cost of the extraction created at *had angered* (as described in (13b)). According to the SSH, this is because, unlike native speakers, L2 learners do not rely on syntactic structures. The relation between *who* and *had angered* is created based on lexico-semantics. Hence, dependency length is the same for (12ab) since in both sentences it is a direct association between the *wh* and the predicate. However, this account is problematic as it is unclear how exactly the *wh* relation is created.

One possibility is so-called *late insertion* according to which the *wh* dependency is only created when the reader notices that an argument is missing. Note that the predicate *is refusing* is the earliest point that makes clear that the direct object of *had angered* is missing; therefore, the *wh* dependency should only be created at *is refusing*. Clearly this is too late to explain the slowdown at *had angered*; that is, late insertion cannot explain the slowdown at *had angered*, as it only predicts a slowdown later at *is refusing*.

The alternative to late insertion is called *active filler strategy* (AFS; Frazier & Clifton, 1989) according to which the *wh*-phrase generates a syntactic requirement in working memory to find an associated gap or empty position (i.e., a position where a constituent such as a direct object is missing). In fact, Clahsen and Felser explicitly assume that L2 readers make use of the AFS (Clahsen & Felser, 2006b). But the *wh* requirement for a related empty position is a syntactic requirement and cannot be characterized as a lexico-semantic constraint. Therefore, in order to explain L2 readers' slowdown at *had angered*,

it is necessary to assume that L2 representations include syntactic relations that span more than just adjacent words.<sup>3</sup>

In sum, the reliable difference described in (13b) cannot be reduced to lexico-semantic and the *wh*-dependency must have been created based on a syntactic representation. The problem then is why the differences in (13ac) were not observed in the L2 data. Note that these are null results, that is, a failure to detect a difference, so in principle they constitute weak evidence as they could be due to lack of statistical power. We know that L2 readers vary more than native speakers in general. In the experiment above, in particular, the L2 participants varied in their comprehension of *wh*-phrases (75% or above according to a pre-test questionnaire) and were not all highly proficient (“at or above the upper intermediate level”, i.e., scored at least 72.5% in a proficiency test; Marinis et al., 2005). Therefore, it is not surprising if the L2 results were noisier and the results weaker compared to native speakers’. To compensate, it might be necessary to collect more data from L2 participants and to include proficiency scores as a factor in the reading-time analyses.

In the number agreement experiment discussed earlier in this chapter, the analyses required proficiency scores so that reliable differences could be detected even though the comparison was between two conditions in simple mono-clausal sentences (see (10) for example sentences). Therefore, it is unsurprising that the much subtler difference in the interaction of the multi-clausal sentences in (13) should also require such scores in their analyses.

As described earlier, spillovers could also increase noise in the L2 data. For example, when comparing *that* in (12a) to *about* in (12b), it is possible that the previous word (or words) had some prolonged effect. It is conceivable that it was more difficult for L2 readers to process *argument* than *argued* and that this difficulty persisted in the following region making it harder to detect the crucial difference.

## **【References】**

- Anderson, J. R. (1974). Verbatim and propositional representation of sentences in immediate and long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 13, 149-162.
- Babaii, E., & Shahri, S. (2010). Psychometric rivalry: The C-test and the close test interacting with test takers’ characteristics. In R. Grotjahn (Ed.), *The C-Test: Contributions from Current Research* (pp. 41-56). Frankfurt: Lang
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Clahsen, H. and Felser, C. (2006a). Grammatical processing in language learners. *Applied*

---

<sup>3</sup>The AFS predicts that readers will attempt to resolve the gap requirement as soon as possible by postulating a gap whenever grammatically permissible. If later words in the sentence disconfirms the position of gap, a slowdown ensues in a *filled-gap effect* (Stowe, 1986, and references therein). The AFS predicts that a gap should be inserted immediately after the verb *argued* in (12a) and a filled-gap effect slowdown should occur at the next word (*that*), which indicates that a gap is not possible at this point. This would suggest that the L1 readers’ slowdown at *that* may be at least in part due to factors other than an intermediate landing position at this point.

- Psycholinguistics*, 27, 3-42.
- Clahsen, H. and Felser, C. (2006b). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, 27, 107-126.
- Clahsen, H., & Felser, C. (2006c). How native-like is non-native language processing? *TRENDS in Cognitive Sciences*, 10, 564-570.
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, 8, e77661.
- Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, 27, 285-319.
- Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4, 93-126.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68, 1-76.
- Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, 7, 55-78.
- Gibson, E., & Wu, H.-H. I. (2013). Processing Chinese relative clauses in context. *Language and Cognitive Processes*, 28, 125-155.
- Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative clause sentences. *Journal of Verbal Learning and Verbal Behavior*, 20, 417-430.
- Hyltenstam, K., & Abrahamsson, N. (2005). Maturation constraints in SLA. In C. J. Doughty & M. H. Long (Eds.) *The Handbook of Second Language Acquisition*. Blackwell Reference Online. 25 February 2017 [http://www.blackwellreference.com/public/book.html?id=g9781405132817\\_9781405132817](http://www.blackwellreference.com/public/book.html?id=g9781405132817_9781405132817)
- Ishizuka, T., Nakatani, K., & Gibson, E. (2006). Processing Japanese relative clauses in context. Paper presented at the nineteenth CUNY conference on human sentence processing, NY, March, 2006.
- Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25, 603-634.
- Jiang, N., Novokshanova, E., Masuda, K. & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, 61, 940-967.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60-99.
- Kim, K. H. S., Relkin, N. R., Lee, K.-M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, 388, 171-174.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580-602.
- Kwon, N., Lee, Y., Gordon, P. C., Kluender, R., & Polinsky, M. (2010). Cognitive and linguistic factors affecting subject/object asymmetry: An eye-tracking study of prenominal relative clauses in Korean. *Language*, 86, 546-582.
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 1-16.
- Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, 47, 50-68.
- Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, 27, 53-78.
- Miyamoto, E. T. (2008). Processing sentences in Japanese. In S. Miyagawa & M. Saito (eds.) *The Oxford Handbook of Japanese Linguistics* (pp. 217-249). Oxford, UK: Oxford University Press.
- Miyamoto, E. T. (2016). Working memory fails to explain subject-extraction advantages (and

- object-extraction advantages) in relative clauses in Japanese. Proceedings of the Japanese Society for Language Sciences 18th Annual International Conference (JLS2016), June 4-5. University of Tokyo, Komaba. pp. 25-28.
- Miyamoto, E. T., & Nakamura, M. (2003). Subject/object asymmetries in the processing of relative clauses in Japanese. In G. Garding & M. Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (pp. 342-355). Somerville, MA: Cascadilla Press.
- Miyamoto, E. T., & Tsujino, K. (2016). Subject relative clauses are easier in Japanese regardless of working memory and expectation. Proceedings of the Japanese Society for Language Sciences 18th Annual International Conference (JLS2016), June 4-5. University of Tokyo, Komaba. pp. 42-45
- Pearlmutter, N.J., Garnsey, S.M. & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language* 41, 427-456.
- R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Schwartz, B.D. & Sprouse, R.A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12, 40-72.
- Sorace, A. (2005). Near-nativeness. In C. J. Doughty & M. H. Long (Eds.) *The Handbook of Second Language Acquisition*. Blackwell Reference Online. 25 February 2017 [http://www.blackwellreference.com/public/book.html?id=g9781405132817\\_9781405132817](http://www.blackwellreference.com/public/book.html?id=g9781405132817_9781405132817)
- Stowe, L. A. (1986). Parsing WH-constructions: evidence for on-line gap location. *Language and Cognitive Processes*, 1, 227-245.
- Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: evidence from eye movements. *Journal of Memory and Language*, 47, 69-90.
- Tsujino, K., & Miyamoto, E. T. (2016). Chinese learners of Japanese read relative clauses faster when subject-extracted in L2 (and object-extracted in L1). Proceedings of the Japanese Society for Language Sciences 18th Annual International Conference (JLS2016), June 4-5. University of Tokyo, Komaba. pp. 50-53
- Ueno, M. & Garnsey, S.M. (2008). An ERP study of the processing of subject and object relative clauses in Japanese. *Language and Cognitive Processes*, 23, 646-688
- Vasishth, S., Chen, Z., Li, Q., Guo, G. (2013). Processing Chinese relative clauses: evidence for the subject-relative advantage. *PLoS ONE*, 8, e77006. doi:10.1371/journal.pone.0077006
- Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- Wagers, M. W., Lau, E.F. & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61, 206-237.
- Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic Theory and Psychological Reality*, 119-161. Cambridge, Mass.: MIT Press.
- Wen, Z., Miyao, M., Takeda, A., Chu, W., & Schwartz, B. D. (2010). Proficiency effects and distance effects in nonnative processing of English number agreement. *Proceedings of the Boston University Conference on Language Development*, 34, 445-456.
- Wilson, B. G., & Miyamoto, E. T. (2015). Proficiency effects in L2 processing of English number agreement across structurally complex material Proceedings of the 151st Meeting of the Linguistic Society of Japan (pp. 176-181). Nagoya University.
- Yamada, T., & Hirose, Y. (2012). Singular/plural asymmetry in Japanese EFL learners' sensitivity to English number dis/agreement. *Proceedings of the Japanese Cognitive Science Society*, 29.