Accepted for publication in Applied Psycholinguistics, July 16, 2018

doi:10.1017/S0142716418000437

Adaptation to syntactic structures in native and non-native sentence comprehension

Edith Kaan*, Corinne Futch

University of Florida

Raquel Fernández Fuertes, Sonja Mujcinovic and Esther Álvarez de la Fuente

University of Valladolid

*Corresponding Author

kaan@ufl.edu

Department of Linguistics

University of Florida

Box 115454

Gainesville, FL 32611

USA

Abstract

Previous research suggests that native speakers quickly adapt to the properties of the language in the surrounding context. For instance, as they repeatedly read a structure that is initially non-preferred or infrequent, they show a reduction of processing difficulty. Adaptation has been accounted for in terms of error-based learning: the error resulting from the difference between the expected and actual input leads to an adjustment of the knowledge representation, which changes future expectations. The present study tested whether experiencing an error is sufficient for adaptation. We compared native English speakers and second-language (L2) learners' processing of, and adaptation to, two types of temporarily ambiguous structures that were resolved towards the non-preferred interpretation. Whereas both native English and L2 speakers showed increased reading times at the disambiguating word versus a non-ambiguous control, our data suggest that only native English speakers adapted, and only to one of the two structures. These results suggest that experiencing an error is not sufficient for adaptation, and that factors such as ease of revision and task effects may play a role as well.

Keywords: syntactic adaptation, second-language processing, filled-gap effect, coordination, sentence processing

Introduction

Recent research suggests that native speakers quickly adapt to the surrounding language context, be it to an interlocutor's accent, vocabulary, or syntactic structures (e.g., Bock, 1986; Brennan & Clark, 1996; Brennan & Hanna, 2009; Kraljic, Samuel, & Brennan, 2008; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). For instance, after repeated exposure to a non-preferred syntactic structure such as a reduced relative clause (e.g. *The soldiers warned about the dangers conducted the midnight raid*), readers show a decrease in processing difficulty for that structure (Farmer, Fine, Yan, Cheimariou, & Jaeger, 2014; Fine & Jaeger, 2013; Fine, Jaeger, Farmer, & Qian, 2013). Readers may even show a preference reversal in that the initially non-preferred structure (e.g. a reduced relative clause) becomes easier to process than the initially preferred structure (a main clause structure in this case, Fine et al., 2013). Syntactic adaptation can also be observed in production as an increased tendency to use a particular structure after this structure has been repeatedly encountered in the recent context (Hartsuiker & Westenberg, 2000; Kaan & Chun, 2018; Kaschak & Borreggine, 2008; Kaschak, Loney, & Borreggine, 2006). Adaptation can be conceptualized as the adjustment of one's linguistic knowledge to accommodate features of the language variety used in the context. Adaptation to a speaker's accent, word choice or syntactic structures may not only facilitate communication, but may also pertain to language learning (Chang, Dell, & Bock, 2006). Given the ubiquity and importance of adaptation, the current study further explores factors driving adaptation by comparing native English speakers and second-language (L2) learners of English.

One mechanism that has been proposed to underlie adaptation is error-based implicit learning (Chang et al., 2006). According to this approach, listeners or readers make implicit predictions regarding upcoming input. When these predictions are not met, the internal knowledge is adjusted, such that future predictions are more likely to be borne out given the context. Evidence for this view is the inverse frequency effect: priming and adaptation effects are stronger for structures that are infrequent, especially for structures that are infrequent given the particular verbs or nouns used, or given other aspects of the context ("surprisal", Bernolet & Hartsuiker, 2010; Hale, 2001; Jaeger & Snider, 2013; Levy, 2008). When an infrequent structure is encountered, the deviance (error) between the actual input and what is expected is larger than when a frequent structure is seen or heard. This larger error will result in a larger adjustment of the knowledge representation. Inverse frequency and surprisal effects have also been observed in children (Peter, Chang, Pine, Blything, & Rowland, 2015), giving support to the idea that error-based learning is a general, life-long learning mechanism for at least some aspects of language (Kaan, 2015).

Alternatively, adaptation can take place through an activation-based mechanism (Reitter, Keller, & Moore, 2011). In this view, structural representations are stored in long-term memory and receive a boost in activation each time they are encountered, leading to a change in the level of resting-state activation. Inverse-frequency effects can be accounted for by assuming that structures that are less frequent have a lower resting-state activation than structures that are more frequent. Exposure to an infrequent structure may lead to a larger boost in activation than exposure to a structure that already is frequent. This leads to larger priming and adaptation effects for less frequent structures. It is not clear how this approach captures finer-grained surprisal effects, however (Fine & Jaeger, 2013).

According to both error-based and activation-based approaches to adaptation, the main factor driving adaptation is frequency. When a sentence is continued in an unexpected, infrequent way, the adjustment of activation of long-term representations will be greater. These approaches predict that, in general, adaptation to an infrequent, non-preferred structure is larger than to its more frequent alternate. Under this approach, one would expect L2 speakers to show larger adaptation effects to infrequent structures than native speakers. This is based on the assumption that L2 and native speakers have different experiences with the alternate structures. First, the relative frequencies with which the alternates are encountered in the L2 learning environment may be different from those experienced by native speakers. It is known that L2 speakers avoid difficult constructions in their own production (Kleinmann, 1978a, 1978b), and have alternates which do not occur in the target variety of the L2, e.g. preposition drop as an alternate to preposition stranding in *wh*-questions (Bardovi-Harlig, 1987; Conroy & Antón-Méndez, 2015; Klein, 1995, 2003). The relative frequency of a less frequent construction compared with a more common alternate may therefore be lower in the language experience of a non-immersed L2 speaker than in that of a native speaker. Second, also the absolute frequency with which the alternates are encountered is different between L2 and native speakers. L2 learners will have had less life-time exposure to L2 structures and their alternates. The effects of frequency on processing are typically logarithmic: a difference in absolute frequency between low-frequency items has a higher impact on language processing than that same absolute difference on the higher end of the frequency scale (Howes & Solomon, 1951). Furthermore, structures that are encountered more often may be more "entrenched" and easier to process than structures that are encountered less frequently in the absolute sense (Schmid, 2007). Differences in absolute frequency between alternate constructions may therefore impact L2 learners more than native speakers, even when the relative frequency between the alternates is comparable for native and L2 exposure. L2 learners may therefore perceive larger differences in frequency between a common syntactic structure and a less common alternate compared with native speakers, and may therefore experience a larger "error", between the initially preferred analysis and the alternate structure, which may lead to stronger adaptation.

However, differences in frequency between the expected and actual structure, and experiencing an "error", may not be sufficient for adaptation. It is likely that factors such as the ease of obtaining the target structure, or the ability to reject the incorrect structure may affect adaptation to the target structure. If this is the case, we can expect that L2 learners will not adapt to non-preferred structures as easily as native speakers. First, adaptation to an initially non-preferred structure may not take place if the reader has difficulties inferring what caused the error and what the target structure is. For instance, if a reader is not able to easily revise *The soldiers warned about the dangers conducted…* into a reduced relative, the reader may not adapt to the reduced relative structure. Some non-preferred continuations of structures are easier to process than others, depending on the semantic and syntactic cues available (Fodor & Inoue, 2000), the relative frequencies of the alternative analyses, or in case of L2 learners, transfer from native language (Dussias & Cramer Scaltz, 2008; Pajak, Fine, Kleinschmidt, & Jaeger, 2016). In addition, L2 learners may experience more difficulty or even break down when reanalyzing sentences due to their processing being less automatic. For instance, Roberts and Felser (2011) report that Greek learners of English successfully revised “easy” garden paths such as *The inspector warned the boss would destroy very many lives*. Here, *the boss* is initially interpreted at the direct object of *warned*, but needs to be reanalyzed as the subject of the embedded clause at *would*. However, the L2 learners broke down in cases such as *While the band played the song pleased all the customers*. Here *the song* is initially taken as the direct object of the verb *played*, but needs to be revised into the subject of the main clause at *pleased.* This revision entails a more drastic change of the thematic structure of the sentence. If the ease of obtaining the target structure affects adaptation, readers, and especially L2 readers, are expected to adapt to a lesser extent to a target structure if it is less frequent or otherwise hard to obtain.

Related to the above, another factor that may affect adaptation is the ability to reject the initial parse. Even if a reader successfully activates a reduced relative structure in *The soldiers warned about the dangers conducted…*, the reader needs to ignore the initial reading according to which the soldiers were the ones giving out the warnings. Native speakers have been shown to keep the initial, incorrect interpretation activated even after successfully obtaining the target structure. For instance, in a study by Kaschak and Glenberg (2004) one group of readers was exposed to the unfamiliar *needs*+Verb construction, such as *the car needs washed*. After exposure, participants showed a decrease in reading times for this construction. However, they also showed facilitation for the modifier construction (*This meal needs washed vegetables*) relative to a participant group that was not trained on the *needs*+Verb construction. This suggests that during exposure also the initially-preferred (modifier) interpretation was activated, and that this structure remained activated above baseline levels even after revision to the target structure (cf. also Christianson, Hollingworth, Halliwell, & Ferreira, 2001). In addition, initial, but ultimately incorrect interpretations have been shown to lead to priming effects even in cases in which readers successfully obtained the target structure (van Gompel, Pickering, Pearson, & Jacob, 2006). Research on L2 learners suggests that learners have more problems rejecting initial, incorrect interpretations (Jacob & Felser, 2016; Pozzan & Trueswell, 2015, 2016), or suffer more from interfering materials in general than do monolinguals (Cunnings, 2016). If the initial structure lingers, this may affect the processing of the next item with a similar ambiguity, which in turn may affect adaptation. If L2 readers indeed have more problems rejecting the initially preferred interpretation than native speakers, adaptation to the non-preferred reading may be weaker in L2 readers.

The factors mentioned above are strongly interrelated, and the present study was not aimed at distinguishing among them. The goal of the present study was to test whether native and L2 speakers differed in their adaptation to non-preferred structures and whether this was affected by the type of structure. We tested adaptation in English speakers and Spanish L2 learners of English to two different syntactic structures that differed in the ease with which the target structure could be obtained, based on frequency, transfer and/or revision cues. If adaptation is mainly based on frequency, adaptation to an infrequent, non-preferred structure should be larger in L2 learners than in native speakers, as the L2 learners have had less absolute and relative exposure to such structures and hence will experience a larger "error" if a sentence no longer continues in the preferred way. However, if the ease of obtaining the target structure and of rejecting the incorrect structure factors in as well, L2 learners are expected to adapt to a lesser extent to non-preferred structures than native speakers, especially for those structures that are harder to obtain.

The present study

In the present study we tested two kinds of non-preferred structures: one in which we expected it to be rather hard to obtain the target reading (filled-gap constructions in *wh*-clauses), and one in which we expected it to be easier to obtain the target reading (*and* coordination ambiguities resolved towards a clausal coordination). The experimental conditions are illustrated in Table 1. Even though the *wh*- and the coordination conditions were presented in the same experiment, they were two different between-item manipulations, and were therefore constructed and analyzed separately.

<<Table 1 around here >>

The first type of experimental conditions were the *wh*-conditions. Examples are given in (1) in Table 1. The intended interpretation of *what* in (1a) is that it is the complement of the stranded preposition *with.* However, when reading the sentence from left to right, readers have a strong tendency to initially interpret *what* in (1a) as the direct object of the verb *repaired*. This holds for native English speakers (e.g. Crain & Fodor, 1985; Omaki et al., 2015; Stowe, 1986) as well as L2 learners of English (e.g. Aldwayan, Fiorentino, & Gabriele, 2010; Dallas, 2008; Felser, Cunnings, Batterham, & Clahsen, 2012; Omaki & Schulz, 2011; Víquez, 2012; Williams, Möbius, & Kim, 2001). This preference can be explained by frequency (*wh*-phrases are more frequently an object than a complement of a stranded preposition, Atkinson & Omaki, 2016), as well as by processing strategies that reduce memory load (Frazier & Clifton, 1989; Gibson, 1998, 2000): assuming that it is costly to have non-integrated information in memory, the parser seeks to integrate a *wh*-phrase as soon as possible. When the direct object position appears to be filled by an overt noun phrase (*the leak* in 1a), an increase in reading times is seen starting at this noun phrase compared with the same noun phrase in a sentence without *what*, as in (1b). This effect is known as the *filled-gap effect* (Crain & Fodor, 1985; Stowe, 1986). The intended interpretation of *what* is that it is the complement of a following preposition (*with* in 1b). However, this intended representation may be relatively hard to obtain or activate at *the leak* in (1a), since there is no information provided by the error as to what the correct analysis is. The only information that can be inferred is that *the leak*, and not *what*, is the direct object of *repair*; no cues are given as to the intended interpretation of *what* (Fodor & Inoue, 2000). If the ease of obtaining the target interpretation affects adaptation, we expected smaller adaptation effects in L2 speakers. We expected that Spanish L2-learners of English, who have been less exposed to this construction in English than native English speakers, and who do not have preposition-stranding in their native language, would initially have difficulty activating this intended interpretation. We therefore expected both native speakers and L2 speakers to show longer reading times at the critical noun for the *what* versus *whether* conditions, replicating other studies; however, we expected that this difference in reading times would decrease in the native speakers as they encountered more filled-gap items such as (1a) in the study, but remain more prominent in the L2 speakers. On the other hand, if frequency differences and experiencing an error are sufficient for adaptation, L2 speakers were expected to show a larger decrease in the size of the disambiguation effect (larger adaptation) than native speakers. Preposition stranding is less expected for L2 speakers, which should lead to a larger error and stronger adaptation.

The second type of experimental condition was the coordination construction illustrated in (2) in Table 1. When reading the *and* condition (2a) from left to right, readers may initially interpret *and* as coordinating the two noun phrases following the verb (*the table and the floor*). The verb *was* is unexpected under this interpretation, leading to an increase in reading time at this verb position in (2a) versus the same verb in the unambiguous control (2b) (Frazier, 1987; Hoeks, Vonk, & Schriefers, 2002). In contrast to the critical noun phrase in the filled-gap constructions (1a), the disambiguating verb *was* in (2a) provides a clear information as to what the correct analysis should be: the verb *was* needs a singular subject, and it is an easy fix to undo the noun-phrase coordination and make the singular noun phrase *the floor* the subject of *was*, retaining the syntactic and thematic structure of the first clause. In addition, Spanish is similar to English in the ambiguity of *and* (*y* in Spanish), in that *and* can coordinate noun phrases as well as clauses. We therefore expected that both native English speakers and L2 learners quickly adapted to this structure: both participant groups were predicted to show a smaller difference in reading times at *was* for (2a) versus (2b) as they encountered more constructions like (2a) in the experiment.

Method

Participants

The native English group consisted of 40 native speakers of American English recruited at the University of Florida, USA (31 women, 9 men; age 18-27, mean age 20.5). The L2 group consisted of 39 Spanish learners of English recruited at the University of Valladolid, Spain (29 women, 9 men, 1 not indicated; age 18-36, mean age 21.5). All participants indicated to have been monolingually raised, to have normal or corrected-to-normal vision, and no dyslexia or reported reading problems. The protocol was approved by the University of Florida Institutional Review Board, and the Ethics Board at the University of Valladolid. Most of the native English participants received course credit for participation; participants at the University of Valladolid received no compensation. Most of the L2 participants indicated to have been learning English since the age of 5, which is common in the contemporary Spanish school and daycare system. Twelve of the L2 speakers indicated to have spent time in an English speaking country. Most of them had spent 2 months or less abroad; one had 9 months of immersion experience, two had 2 years and one participant had 3 years of English immersion experience. The L2 participants indicated to be currently using English 32% percent of the time (*SD* 13.5) and rated their own English speaking, listening, and reading proficiency as 7.1 (*SD* 1.0), 7.9 (*SD* 1.0), and 8.1 (*SD* 1.0) out of 10 respectively, where 10 is native proficiency. All participants completed the LexTALE task (Lemhöfer & Broersma, 2012). Performance on this lexical decision task highly correlates with other language proficiency measures (Lemhöfer & Broersma, 2012). The native English group had a significantly higher score on the LexTALE task than the L2 group (Native English: mean 92.2, range 71.25-100; L2: mean 72.4, range 55-100, $T(77) = 9.68$, $p < .0001$). Of the native English speakers, 38 scored within the advanced range on the LexTALE task (score of 80-100), and two in the intermediate range (60-80). Of the L2 group, nine scored in the advanced range, 27 in the intermediate, and three in the low range (score of 59 or lower).

Stimuli

Thirty-six pairs of sentences were constructed of type (1) in Table 1 (*wh*-conditions), and another 36 pairs of type (2) in Table 1 (coordination conditions). In the *wh*-conditions, the question word (*what*) was intended as the complement of a preposition. The verb in the embedded clause was always followed by a noun phrase in order to elicit the filled-gap effect. In the control condition, *whether* replaced *what*, and a noun phrase followed the preposition. The critical position was therefore the noun phrase after the embedded verb (underscored in Table 1 for the purpose of illustration). Here we expected an increase in reading times for the *what* versus *whether* condition, especially at the start of the task. In the coordination conditions, the two noun phrases were separated by *and* in the *and* coordination condition (2a) and by *but* in the control condition (2b). The latter is not very likely to indicate a coordination between two noun phrases. The critical word was the verb (*was*) following the second noun phrase, signaling that the second noun was the subject of a new clause, and that the correct analysis in (2a) was one in which *and* coordinates two clauses rather than two noun phrases. In all experimental items, the critical word was in the sixth presentation frame, followed by at least three segments. Experimental items were divided into two counterbalanced presentation lists, such that a participant only saw one version of each experimental item pair, and each list contained 18 different items for each of the four experimental conditions. Within and across lists, the items in the *what* and *whether* condition were matched in the length in number of characters, and word form frequency as determined by the Corpus of Contemporary American English (Davies, 2008-) of the verb preceding the critical noun phrase, and of the noun in the critical position; items in the *and* condition were matched with those in the *but* condition on the length and frequency of the noun preceding the critical verb.

In addition to the experimental items, we constructed 72 distractor sentences that had other syntactic structures. Similar to the experimental items, the distractor sentences also consisted of a main and an embedded clause. The embedded clause was either the complement of the main clause verb (e.g. *The pilot saw that the weather was too stormy for the plane to take off*), and started with *that* (18 items), *when*, *how*, or *why* (6 each), or was an adjunct clause (e.g. *The baby played on the blanket while the grandmother sipped her coffee*). Adjunct clauses either followed (18 items) or preceded the main clause (18 items). To avoid unintended effects on adaptation, distractor materials never contained *and*, *but*, *what* or *whether*, other noun phrase coordinations, or sentences with stranded prepositions. A complete list of materials is included as supplementary materials. The order of the 72 experimental items and 72 distractors was automatically pseudorandomized for each participant, such that items from the same experimental main type (*wh*, coordination) were separated by at least one distractor item or experimental item of a different main type.

To encourage participants to pay attention to the sentences, two-thirds of the experimental items and one-third of the distractors (50% of all items in total) were followed by a yes/no comprehension question. The correct answer was 'no' in half of the questions and 'yes' in the other half. Questions mainly probed which antagonist did what. We did not systematically probe incorrect or target interpretations of the ambiguous structures; however, the coordination conditions had seven comprehension questions that probed the lingering of the initial reading. For instance, the sentence *The dog buried the bone and the stick was left behind the doghouse* was followed by the question *Did the dog bury the stick?* A 'yes' answer to this question suggests that the reader still entertained the reading in which *the bone and the stick* are both direct objects of the verb *bury*. We will discuss performance on these questions separately in the results section.

Procedure

Each participant was randomly assigned to only one presentation list. Sentences were presented in a non-cumulative moving window self-paced reading paradigm controlled by Linger (developed by Doug Rohde, http://tedlab.mit.edu/~dr/Linger/). Each trial started with a sentence contour: all words and spaces were replaced by dashes. Participants controlled the presentation of the words using the space bar. Each time they pressed the space bar, a new word was presented, and the previous word was replaced with dashes. At the end of the sentence, the next trial was displayed, or a comprehension question was presented. Participants answered the question by pressing the ‘f’ or ‘j’ key, corresponding to the letters *Y* and *N* displayed at the left and right side on the screen, respectively. Participants were instructed to read at a normal pace. They received five practice items (two followed by questions) that contained a main clause and an embedded clause, but otherwise did not resemble the experimental items. A short break was automatically enforced in the middle of the self-paced reading experiment. The self-paced reading task was followed by the LexTALE lexical decision task (Lemhöfer & Broersma, 2012), and a language background questionnaire.[1]

Analysis and results

Reading times

Analysis. Following experimental conventions, we first omitted reading times that were too short or too long to reflect reading processes, and then transformed the data to approximate a normal distribution (Baayen & Milin, 2010). We omitted data points longer than 5000 ms and shorter than 100 ms. This procedure affected less than 0.01% of all data points in either group. Next, we omitted data points that were longer than the mean plus 2.5 SD for each participant. This affected 2.5% and 2.9% of the data points in total for the native English and L2 groups respectively. Restricted to the critical word positions of our experimental conditions, these cutoff procedures affected, in the *what* condition, 5.8% for the native English, and 3.4% for the L2 group; in the *whether* condition 3.3% and 2.8%; in the *and* condition 1.5% and 0.6%; and in the *but* condition 0.7% and 0.6%. We then log-transformed the reading times (natural logarithm) to adjust for the skewedness of the distribution. The Box-Cox procedure (Box & Cox, 1964) confirmed that a log transformation was appropriate (smallest λ was -0.6). For the native and L2 groups separately, we calculated residual reading times based on a linear mixed effects model on all data (experimental items as well as distractors), with the length of the word in the number of characters, and the (natural) log-transformed position of the trial in the experiment as fixed effects. Random effects included by-participant intercepts and by-participant slopes for word length and the log of the trial position. We included trial position as a factor to control for overall effects of the duration of the experiment, regardless of the distribution of the experimental conditions. The main analyses were conducted on the residual log reading times thus obtained. Word length, the overall position of the trial in the experiment, and overall reading speed of each participant are all strong predictors of reading times, and it is standard in analyses of reading times to use residual rather than raw reading times to reduce these effects.[2] Since the two experimental manipulations (*wh*- and coordination conditions) involved different lexical items and different constructions, they were not directly comparable. We therefore analyzed these sentence types separately. For each main type, we analyzed the residual log reading times at the critical position using a linear mixed effects model in R, version 3.4.3 (R Core Team, 2015), using the lme4 package, version 1.1-15 (Bates, Maechler, Bolker, & Walker, 2015). Fixed effects were Language Group (deviation coded, with L2 coded +0.5), Condition (Ambiguous/Non-ambiguous, deviation coded with Ambiguous coded +0.5), and the Number of preceding temporarily ambiguous items seen of the type under investigation (centered). Previous studies (Fine et al., 2013; Kaschak & Glenberg, 2004) suggest that adaptation can take place within a few trials. To better capture the early part of the study, we used a (natural) log-transformation of the number of preceding ambiguous structures seen.[3] We augmented the number by 1 before transformation to avoid taking the log of 0. Results were not qualitatively different when a non-transformed number was used. We first estimated models with a maximal random effects structure. When these models did not converge, we removed the correlations between the random slopes and intercepts (Barr, Levy, Scheepers, & Tily, 2013).[4] We first conducted an analysis with native and L2 speakers combined. Since the main aim of the study was to explore to what extent native and L2 speakers adapt, we also analyzed the participant groups separately. The analysis of the L2 data included as fixed effects the factor proficiency (LexTALE score) and its interactions with Condition and the Number of *what* or *and* items seen. *P*-values were estimated using lmerTest, version 2.0-36 (Kuznetsova, Brockhoff, & Christensen, 2016), which bases degrees of freedom on Satterthwaite approximations.

Wh-conditions. Mean residual log reading times for the *what* and *whether* sentences for the native English and L2 groups are given in Figure 1. Replicating prior studies, reading times were longer at the noun phrase following the verb (position 6) in the *what* than in the *whether* conditions, when it became clear that *what* could no longer serve as the direct object of the verb (e.g. Aldwayan et al., 2010; Crain & Fodor, 1985; Dallas, 2008; Felser et al., 2012; Omaki et al., 2015; Omaki & Schulz, 2011; Stowe, 1986; Víquez, 2012; Williams et al., 2001).

<<Figure 1 around here >>

Figure 2 shows the change of the reading times at the critical noun phrase (position 6 in Table 1) as a function of the number of *what* constructions encountered.

<<Figure 2 around here >>

Results from the linear mixed effects model on the reading times at the post-verbal noun phrase, comparing the two participant groups, are given in Table 2. For all participants taken together, the *what* condition was read more slowly than the *whether* condition (effect of Condition, Table 2); however, there was no effect of adaptation: the difference between *what* and *whether* did not decrease as more *what* conditions had been encountered (no interaction of Condition by Number of preceding *what* sentences seen).

<<Table 2 around here >>

Even though no interaction with Language Group was significant, we nevertheless conducted separate analyses for each of the two participant groups, to see if both participant groups showed the same pattern of effects. Table 3 lists the results for the native English group; Table 4 lists the results for the L2 group. The difference in reading times between the *what* and *whether* conditions at the critical position failed to reach significance for the native English group. The L2 group showed significantly longer reading times at the direct object for the *what* vs. *whether* condition, suggesting they experienced difficulty when the gap-position was filled. However, the difference in reading time between the conditions was not affected by the Number of *what* conditions seen. L2 proficiency as measured by the LexTale score had no effect.

<<Table 3 around here >>

<< Table 4 around here >>

Coordination conditions. The mean reading times for the coordination conditions are given in Figure 3. As expected both participant groups showed a longer reading time at the disambiguating finite verb in the ambiguous *and* condition versus the *but* control condition. This suggest that both participant groups had a preference for a noun phrase coordination and experienced processing difficulty when this analysis was no longer compatible with the incoming information.

<<Figure 3 around here >>

Results from the linear mixed effects model on the reading times for the disambiguating verb are given in Table 5. Critically, there was a main effect of Condition, and a three-way interaction between Language Group, Condition, and Number of preceding *and* sentences seen.

<<Figure 4 around here >>

<< Table 5 around here >>

Separate analyses for the native and L2 groups for the disambiguating position (position 6) are given in Tables 6 and 7. The native English group showed a significant interaction of Condition by Number of preceding *and* sentences: the difference in reading times between the *and* and *but* conditions became smaller as more *and* structures had been seen (see Figure 4a). The L2 learners, on the other hand, showed a numerically *larger* difference between the two conditions as they had read more items with *and* (Figure 4b); however, there was no interaction between Condition and the Number of *and* sentences seen (Table 7). Again, we did not see any effect of proficiency as measured by the LexTale score. In sum, the results suggest that the significant three-way interaction in the overall analysis is due to the fact that the native English group shows a significant two-way interaction between condition and trial number, while this two-way interaction is absent in the L2 group.

<<Table 6 around here >>

<<Table 7 around here >>

Comprehension questions

Two-thirds of the experimental items were followed by a comprehension question. We used the accuracy on the questions to (1) probe further differences between the groups; and more specifically, (2), to explore whether the groups differed in the lingering of the initial interpretation of the ambiguous *and* sentences.

Differences between native and L2 speakers. Mean accuracy to the comprehension questions in the study is given in Table 8. The lower performance in the *whether* condition for both groups can be due to the ambiguity of some of the questions in this condition. For instance, the question *Did the worker repair the leak?* has no obvious correct answer following *The builder wondered whether the worker repaired the leak with some tape before going home*. The *whether* condition was therefore dropped from further analysis of the question data.

<<Table 8 around here >>

We conducted logistic linear mixed effects analyses on the response accuracy of the questions following the experimental items as a function of Condition and Language Group. We analyzed the *what* condition and the coordination conditions separately. The L2 group responded less accurately than the native group in both the *what* (Estimate: -0.61; *S.E.* 0.25; *z*-value -2.46; $p < .05$) and the coordination conditions (Estimate -0.34; *S.E.* 0.17; *z*-value -2.03; $p < .05$). No effects of Condition were found, suggesting that overall performance on the questions was not affected by ambiguity.

Lingering of the initial interpretation. Although the current experiment was not designed to systematically test whether readers had difficulty rejecting the initial interpretation of the ambiguous structures, the coordination condition had seven comprehension questions that probed the lingering of the initial reading. This is the interpretation in which the noun phrase after *and* was the object of the preceding verb. Mean accuracy for these questions in the native English group was 0.80 (*SD* 0.30) for *and*, and 0.92 (0.15) for *but*. Mean accuracy on these questions in the L2 group was 0.82 (0.20) for *and*, and 0.85 (0.19) for *but*. A paired t-test on the ambiguous versus unambiguous conditions suggests that the native English speakers performed worse on the questions probing the lingering interpretation in the ambiguous *and* versus unambiguous *but* condition, $T(39) = -2.74$, $p < .01$, but that the L2 group showed no difference, $T(38) = -0.80$, $p = .43$. Additionally, six questions in the coordination conditions probed the target interpretation of the noun phrase after *and* (e.g. *The boy rolled up the carpet and the rug was moved by the girl. Did the girl move the rug?* Correct answer 'yes'). Both groups responded to the same level of accuracy, with no difference between the *and* and the *but* conditions (L2 group: mean accuracy *and*: 0.90, *SD* 0.18; *but*: 0.90, *SD* 0.18; $T < 1$; native English group: mean accuracy *and*: 0.92, *SD* 0.16; *but*: 0.93, *SD* 0.12; $T < 1$). We therefore have no evidence that the L2 group had specific difficulties with obtaining or reconstructing the correct interpretation in the ambiguous *and* condition, or with rejecting the initial interpretation in this condition.

Discussion

The aim of the present study was to test syntactic adaptation in native and L2 speakers. We presented native English and Spanish L2 English readers with two different syntactic structures that differed in the ease with which the intended non-preferred reading could be obtained: coordination and *wh*-ambiguities. If frequency differences and experiencing an "error" (that is, a deviation from an expected structure) are sufficient for adaptation, adaptation to an infrequent, non-preferred structure should, in general, be larger in L2 learners than in native speakers: L2 learners have had less relative and absolute exposure to such non-preferred structures, and would therefore experience a larger "error" when the structure continues as the non-preferred alternate, resulting in a larger adjustment to this non-preferred alternative over the course of the experiment. However if, in addition or instead, the ease of obtaining the target structure or the lingering of an incorrect interpretation affects adaptation, we expected L2 speakers to show weaker adaptation than the native speakers, especially for the *what* conditions which did not have a structural equivalent in Spanish, and in which the disambiguating position did not contain any direct cues as to the intended, target representation.

Our results suggest that frequency differences between the typical and target structure, and experiencing an "error" are not sufficient for adaptation. For both coordination and *wh*-constructions, longer reading times were observed at the critical, disambiguating position. This suggests that both groups had a preference for interpreting *what* as the direct object of the verb, and for taking *and* as a coordination of two noun phrases, and that both groups experienced processing difficulty ("error") when the preferred, expected analyses could no longer be pursued. However, we found evidence of adaptation only in the native English group and only to the *and* coordination condition: the more *and* items had been encountered that were resolved towards the initially non-preferred clausal coordination, the smaller the difference in reading times at the point of disambiguation versus the unambiguous *but* control sentences. We do not have evidence for adaptation to either the *and* or *what* condition in the L2 group. We also did not find adaptation to the *what* condition in the native English speakers, at least not in the analysis reported in the main text. The latter finding is rather surprising given previous reports of adaptation to complex constructions in native speakers. Of course, we may just not have had enough power, or may not have used a long enough study to obtain adaptation effects in these cases. Below we will discuss other potential reasons why we did not observe adaptation effects in our *what* conditions, and why L2 speakers may not have adapted to the *and* conditions whereas our native speakers did.

Failure to find adaptation to the what *conditions in native speakers*

We did not find adaptation effects for the filled-gap (*what*) conditions, not even in our native speakers. This is in contrast to previous studies reporting decreases in garden path effects in rather complex structures such as reduced relatives and object relatives (Farmer et al., 2014; Fine & Jaeger, 2013; Fine et al., 2013; Wells et al., 2009), but see Stack, James, and Watson (2018) for a recent failure to replicate adaptation effects. The difference in outcomes between the current and prior studies may be due to differences in the design of the experiment, and in the way outliers in the data are treated. Most studies reporting adaptation have not used many distractor items, or none at all during an exposure phase (e.g. experiment 2 in Fine et al., 2013). Myslín and Levy (2016) report that adaptation is stronger when critical structures are blocked. The fact that in our study the *what* items were interleaved with distractor items and coordination conditions and consisted only of 12% of the items in the study, may therefore have hindered adaptation especially of a complex construction such as our *what* conditions.

Another difference between previous studies and the current study is the treatment of outliers in the data. Previous studies typically only omitted response times longer than 2000 ms (e.g. Fine et al., 2013), without removing outliers on a by-participant basis like we did in the analyses reported above. When we analyzed our data using cutoff criteria similar to those used in previous studies, the adaptation effect for the *what* vs. *whether* condition was significant in the native English data, but not in the L2 data (see supplementary materials C). The cutoff procedure used in the analysis reported in the main text above resulted in a loss of 5% of the native data at the filled-gap position versus 0.5% in the analysis using a liberal cutoff procedure. The adaptation effect observed in the latter analysis was, therefore, driven by only a small number of extreme data points. It is, therefore, possible that adaptation effects reported in prior studies using less stringent cutoff procedures were driven by outliers and thus were not very robust.

Ease of obtaining the target structure

The observation that our native speakers adapted more robustly to the *and*-coordination condition than to the *what* condition needs explanation. As discussed in the introduction, it may have been harder to obtain the target structure in the filled-gap (*what*) conditions than in the coordination conditions. In the filled-gap condition the presence of an overt object noun phrase after the verb indicates that the preferred interpretation (*what* as the direct object) is incorrect. However the noun phrase carries no cues as to what the intended structure is; on the other hand, the disambiguating verb in the coordination condition does provide cues as to how the initial coordination between two direct objects would need to be revised (Fodor & Inoue, 2000). Readers may, therefore, be less likely to adapt to the target structure if it is harder to infer the target structure at the point of the error.

L2 speakers, being less familiar with the target structures, may have had more difficulty obtaining the target interpretation, and may have therefore been even less likely to adapt to the target structures; this in spite of the fact that they noticed at the point of disambiguation that the preferred analysis could no longer be pursued, and that they were eventually successful in obtaining or reconstructing the target structure when answering the comprehension questions (our L2 learners performed with 90% accuracy on questions probing the target interpretation in the coordination conditions). The fact that our L2 speakers showed effects of disambiguation at the same word position as the native speakers suggests that the lack of adaptation in the L2 group cannot be due to the L2 speakers delaying their processing (Dekydtspotter, Schwartz, & Sprouse, 2006), assigning only a shallow representation to the sentence (Clahsen & Felser, 2006), or not predicting (Hopp, 2015; Kaan, Kirkham, & Wijnen, 2016; Lew-Williams & Fernald, 2010; Martin et al., 2013).

Note that our findings are not incompatible with an error-based learning approach. If the intended parse cannot be easily identified, the deviance between the initially pursued structure and the target structure cannot be easily calculated, leading to a noisier error signal, which in turn leads to weaker adaptation. Assuming that error signals experienced by L2 processers are more inconsistent or noisier than in native speakers, the failure for the L2 speakers to adapt even with respect to the *and* conditions can be accounted for[5].

Lingering initial interpretations

In the introduction we mentioned the lingering activation of the initial, incorrect reading as another potential factor affecting adaptation. After encountering the critical noun phrase (filled-gap) in the *wh*-conditions, or the critical verb in the *and* coordinations, the initial analysis must be rejected and a new analysis activated or built. The initial analysis may however linger and lead to priming of the direct object interpretation of a fronted *what* phrase, and of the noun phrase coordination (cf. van Gompel et al., 2006 for priming by lingering interpretations). This may in turn hinder adaptation to the intended structures. Results from the small number of questions that probed the lingering representation in the coordination condition suggest that, if anything, our native speakers had more difficulty rejecting the initial *and* coordination interpretation than our L2 speakers. The L2 speakers' overall worse performance on the questions, and the fact that their response accuracy was not different to critical questions following *and* compared to *but* sentences, suggests that they may have problems with interfering information in general, not specifically restricted to interference of lingering readings (Cunnings, 2016). Adaptation is, therefore, probably not much affected by the lingering of the initial, incorrect, interpretation.

Task demands and other factors

The ease of obtaining the intended target interpretation is likely not the only factor affecting adaptation. Numerically, the L2 speakers showed a trend towards an anti-adaptation effect in the coordination conditions: the difference in reading times between the non-preferred and preferred condition became numerically larger as more *and* coordinations had been encountered. Although speculative at this point, we suggest that also task demands may affect adaptation. In our study, more than half of the questions following the coordination sentences explicitly probed the noun phrase following *and*. Participants may have noticed this and may therefore have started paying more attention to the point of disambiguation (Swets, Desmet, Clifton, & Ferreira, 2008), leading to longer reading times at the critical position in this condition as the experiment progressed. Recall that the native speakers rapidly adapted to the clausal *and* coordinations. The difference between the L2 and native speakers could be attributed to a difference in sensitivity to task demands between the two groups. The L2 group may have felt more pressured to do well on the task than the native group. L2 speakers may have processed the sentences more strategically, counteracting implicit learning effects that would give rise to adaptation (Kaschak & Glenberg, 2004; Traxler & Tooley, 2008). This explanation can be tested in future experiments in which task demands are varied.

We also like to point out that our L2 and native groups differed in the testing environment, school system, and many other factors that were beyond our control, which may also have contributed to the differences found between the groups, and which may have made the groups hard to compare (Dekydtspotter et al., 2006). To further explore what factors affect adaptation, future studies should therefore test L2 groups with similar language and educational backgrounds and in the same location, but with, e.g., a wide range of proficiency levels.

Conclusion

Our results suggest that differences in frequency and experiencing an error, are not sufficient to adapt to a non-preferred syntactic structure: both native and L2 speakers showed garden path effects, but only native speakers showed adaptation effects and only in one of the conditions. Adaptation may therefore be affected by various factors, such as the ease of obtaining the target structure and task demands. Future research should be directed towards further identifying these factors, investigating how adaptation can be boosted, and exploring the relation between adaptation and longer-term retention.

Acknowledgements

The authors would like to thank Patricia Aziz for the preparation of the stimulus materials, Andrea Dallas for feedback on the project, and Devin McCann and Sara Levy for their help collecting data. This research was supported in part by a 2016-2017 University of Florida University Scholars Program awarded to CF, and by the Castile and León Regional Government Education Division (Spain) and FEDER [VA009P17].

Notes

1. The native speakers also completed a digit span task after the self-paced reading task. Due to practical constraints, we could not do this task in the L2 group; we therefore did not further consider digit span in the analysis of the data.

2. Analyses on the non-transformed, raw reading times with a cutoff similar to that reported in the main text yielded qualitatively similar results for the coordination condition as those reported in the main text, except that there was no interaction between Language Group, Condition and Number of preceding *and* items. Note that raw reading times are not corrected for overall differences in reading speed between participants. The large overall difference in raw reading times between the native English and the L2 speakers may therefore have made the triple interaction harder to obtain. In contrast to the analysis reported in the main text, the analysis of the *wh*-conditions showed a significant interaction of Condition × Number of preceding *what* items seen, which was mainly driven by the native English speakers; L2 speakers did not show this interaction in the by-group analysis. Since data were not log-transformed, the adaptation effect in the native English speakers may have been driven by a few long RTs. The log-transformation used in the analysis in the main text made these data points less influential. Figures of raw reading time patterns and tables with results from the statistical analyses on the raw reading times are given in the supplementary materials.

3. We included only the number of temporarily ambiguous structures seen rather than collapsing over the two experimental conditions per type. This was motivated by the idea that the *what*–condition had a different syntactic structure than the *whether* control condition. Analyses in which we replaced the number of temporarily ambiguous conditions seen with the number of unambiguous conditions (*whether* or *but*) seen yielded no significant adaptation effects in the

wh-conditions. In the coordination conditions, replacing the number of *and* items seen with the number of *but* items seen yielded effects similar to those reported for the main text for the overall analysis. This is not surprising given that the *and* and the *but* conditions have the same structure.

4. Since we were interested in the effects of Condition, the Number of preceding structures seen, and their interactions, including all these effects as random slopes was theoretically motivated. However, we are aware that this may have led to overparametrization and a reduction of power (Bates, Kliegl, Vasishth, & Baayen, 2015). However, fixed effects that were significant in the maximal random effects models were also significant in models with only random intercepts, and vice-versa; the only exception being the model for native English for the *wh*-conditions. With a minimal random effects-structure, the effect of Condition and the effect of the Number of *what* items seen were significant, whereas these effects were not significant with a maximum model.

5. We like to thank one of the reviewers for pointing this out.

References

Aldwayan, S., Fiorentino, R., & Gabriele, A. (2010). Evidence of syntactic constraints in the processing of wh-movement. A study of Najdi Arabic learners of English. In B. VanPatten & J. Jegerski (Eds.), *Research in second language reading and parsing* (pp. 65-86). Amsterdam/Philadelphia: John Benjamins.

Atkinson, E., & Omaki, A. (2016). *Adaptation of gap predictions in filler-gap dependency processing*. Paper presented at the 29th Annual CUNY Conference on Human Sentence Processing, Gainesville, FL.

Baayen, H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3, 12-28.

Bardovi-Harlig, K. (1987). Markedness and salience in second-language acquisition. *Language Learning*, 37, 385-407. doi:10.1111/j.1467-1770.1987.tb00577.x

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory & Language*, 68, 255-278. doi:10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *ArXiv e-print*. Retrieved from http://arxiv.org/abs/1506.04967 (ArXiv e-print)

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1-48. doi:10.18637/jss.v067.i01

Bernolet, S., & Hartsuiker, R. J. (2010). Does verb bias modulate syntactic priming? *Cognition*, 114, 455-461. doi:http://dx.doi.org/10.1016/j.cognition.2009.11.005

Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355-387. doi:http://dx.doi.org/10.1016/0010-0285(86)90004-6

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B (Methodological)*, 26, 211-252.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1482-1493. doi:10.1037/0278-7393.22.6.1482

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, 1, 274-291. doi:10.1111/j.1756-8765.2009.01019.x

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113, 234-272. doi:10.1037/0033-295x.113.2.234

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42, 368-407. doi:http://dx.doi.org/10.1006/cogp.2001.0752

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

Conroy, M. A., & Antón-Méndez, I. (2015). A preposition is something you can end a sentence with: Learning English stranded prepositions through structural priming. *Second Language Research*, 31, 211-237. doi:10.1177/0267658314555945

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 94-128). Cambridge: Cambridge University Press.

Cunnings, I. (2016). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, 1-20.

Dallas, A. C. (2008). *Influences of verbal properties on second-language filler-gap resolution: A cross-methodological study.* (Ph.D.), University of Florida.

Davies, M. (2008-). The Corpus of Contemporary American English (COCA): 520 million words, 1990-present. Available online at http://corpus.byu.edu/coca/.

Dekydtspotter, L., Schwartz, B. D., & Sprouse, R. A. (2006). The comparative fallacy in L2 processing research. In M. G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp. 33-40). Somerville, MA: Cascadilla Proceedings Project.

Dussias, P. E., & Cramer Scaltz, T. R. (2008). Spanish-English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologia*, 128, 501-513.

Farmer, T. A., Fine, A. B., Yan, S., Cheimariou, S., & Jaeger, T. F. (2014). Error-driven learning of higher-level expectancies during reading. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 2181-2186). Austin, TX: Cognitive Science Society.

Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, 34, 67-98. doi:10.1017/S0272263111000507

Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, 37, 578-591. doi:10.1111/cogs.12022

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, 8, 1-18. doi:10.1371/journal.pone.0077661

Fodor, J. D., & Inoue, A. (2000). Syntactic features in reanalysis: Positive and negative symptoms. *Journal of Psycholinguistic Research*, 29, 25-36. doi:10.1023/A:1005168206061

Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language & Linguistic Theory*, 5, 519-559.

Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4, 93-126. doi:10.1080/01690968908406359

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68, 1-76. doi:http://dx.doi.org/10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language, Brain* (pp. 95-126). Cambridge, MA: MIT Press.

Hale, J. (2001). *A probabilistic Earley parser as a psycholinguistic model.* Paper presented at the Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies.

Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition,* 75, B27-B39. doi:http://dx.doi.org/10.1016/S0010-0277(99)00080-3

Hoeks, J. C. J., Vonk, W., & Schriefers, H. (2002). Processing coordinated structures in context: The effect of topic-structure on ambiguity resolution. *Journal of Memory and Language,* 46, 99-119. doi:http://dx.doi.org/10.1006/jmla.2001.2800

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching,* 53, 277-306. doi: 10.1515/iral-2015-0014

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word probability. *Journal of Experimental Psychology,* 41, 401-410.

Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *The Quarterly Journal of Experimental Psychology,* 69, 907-925. doi:10.1080/17470218.2014.984231

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition,* 127, 57-83. doi:http://dx.doi.org/10.1016/j.cognition.2012.10.013

Kaan, E. (2015). Knowing without predicting, predicting without learning. *Linguistic Approaches to Bilingualism*, 5, 482–486. doi:10.1075/lab.5.4.07kaa

Kaan, E., & Chun, E. (2018). Priming and adaptation in native speakers and second-language learners. *Bilingualism: Language and Cognition*, 21, 228−242. doi:https://doi.org/10.1017/S1366728916001231

Kaan, E., Kirkham, J., & Wijnen, F. (2016). Prediction and integration in native and second-language processing of elliptical structures. *Bilingualism: Language and Cognition*, 19, 1-18. doi:10.1017/S1366728914000844

Kaschak, M. P., & Borreggine, K. L. (2008). Is long-term structural priming affected by patterns of experience with individual verbs? *Journal of Memory and Language*, 58, 862-878. doi:10.1016/j.jml.2006.12.002

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, 133, 450–467. doi:10.1037/0096-3445.133.3.450

Kaschak, M. P., Loney, R. A., & Borreggine, K. L. (2006). Recent experience affects the strength of structural priming. *Cognition*, 99, B73-B82. doi:http://dx.doi.org/10.1016/j.cognition.2005.07.002

Klein, E. C. (1995). Evidence of a wild grammar. When PPs rear their empty heads. *Applied Linguistics*, 16, 87-117.

Klein, E. C. (2003). *Toward second language acquisition. A study of null-prep*. Dordrecht: Kluwer.

Kleinmann, H. H. (1978a). Avoidance behavior in adult second language acquisition. *Language Learning*, 27, 93–107.

Kleinmann, H. H. (1978b). The strategy of avoidance in adult second language acquisition. In W. C. Ritchie (Ed.), *Second language acquisition research: Issues and implications* (pp. 157–174). New York: Academic Press.

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science,* 19, 332-338.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models (Version R package version 2.0-33). Retrieved from https://CRAN.R-project.org/package=lmerTest

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavioral Research*, 44, 325-343. doi:10.3758/s13428-011-0146-0

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126-1177. doi:http://dx.doi.org/10.1016/j.cognition.2007.05.006

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language,* 63, 447-464. doi:10.1016/j.jml.2010.07.003

Martin, C., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, 69, 574-588. doi:http://dx.doi.org/10.1016/j.jml.2013.08.001

Myslín, M., & Levy, R. (2016). Comprehension priming as rational expectation for repetition: Evidence from syntactic processing. *Cognition*, 147, 29-56. doi:http://dx.doi.org/10.1016/j.cognition.2015.10.021

Omaki, A., Lau, E. F., Davidson White, I., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology*, 6. doi:10.3389/fpsyg.2015.00384

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constaints in second-language sentence processing. *Studies in Second Language Acquisition*, 33, 563-588. doi:10.1017/S0272263111000313

Pajak, B., Fine, A. B., Kleinschmidt, D. F., & Jaeger, T. F. (2016). Learning Additional Languages as Hierarchical Probabilistic Inference: Insights From First Language Processing. *Language Learning*, 66, 900-944. doi:10.1111/lang.12168

Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language*, 81, 1-15. doi:http://dx.doi.org/10.1016/j.jml.2014.12.002

Pozzan, L., & Trueswell, J. C. (2015). Revise and resubmit: How real-time parsing limitations influence grammar acquisition. *Cognitive Psychology*, 80, 73-108. doi:http://dx.doi.org/10.1016/j.cogpsych.2015.03.004

Pozzan, L., & Trueswell, J. C. (2016). Second language processing and revision of garden-path sentences: a visual word study. *Bilingualism: Language and Cognition*, 19, 636–643. doi:10.1017/S1366728915000838

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive Science*, 35, 587-637. doi:10.1111/j.1551-6709.2010.01165.x

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, 32, 299-331. doi:10.1017/S0142716410000421

Schmid, H.-J. (2007). Entrenchment, salience, and basic levels. In D. Geeraerts, H. Cuyckens, & H.-J. Schmid (Eds.), *The Oxford Handbook of Cognitive Linguistics* (pp. 117-138). Oxford: Oxford University Press.

Stack, C. M. H., James, A. N., & Watson, D. G. (2018). A failure to replicate rapid syntactic adaptation in comprehension. *Memory & Cognition*. doi:10.3758/s13421-018-0808-6

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes,* 1, 227-245. doi:10.1080/01690968608407062

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition,* 36, 201-216. doi:10.3758/mc.36.1.201

Traxler, M., & Tooley, K. (2008). Priming in sentence comprehension: Strategic or syntactic? *Language and Cognitive Processes,* 23, 609-645. doi:10.1080/01690960701639898

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language,* 55, 335-362. doi:http://dx.doi.org/10.1016/j.jml.2006.06.004

Víquez, A. J. C. (2012). *Online processing of wh-dependencies in English by native speakers of Spanish.* (Ph.D.), University of Kansas.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology,* 58, 250-271. doi:http://dx.doi.org/10.1016/j.cogpsych.2008.08.002

Williams, J. N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: parsing strategies and plausibility constraints. *Applied Psycholinguistics,* 22, 509–540.

Table 1. Experimental conditions

Wh-conditions	
1a. *What* (ambiguous)	$\vert_1$ The builder $\vert_2$ wondered $\vert_3$ *what* $\vert_4$ the worker $\vert_5$ repaired $\vert_6$ **the leak** $\vert_7$ with $\vert_8$ before $\vert_9$ going home. $\vert$
1b. *Whether* (control)	$\vert_1$ The builder $\vert_2$ wondered $\vert_3$ *whether* $\vert_4$ the worker $\vert_5$ repaired $\vert_6$ **the leak** $\vert_7$ with $\vert_8$ some $\vert_9$ tape $\vert_{10}$ before $\vert_{11}$ going home. $\vert$
Coordination conditions	
2a. *And* (ambiguous)	$\vert_1$ The servant $\vert_2$ cleaned $\vert_3$ the table $\vert_4$ *and* $\vert_5$ the floor $\vert_6$ **was** $\vert_7$ cleaned $\vert_8$ by $\vert_9$ the maid. $\vert$
2b. *But* (control)	$\vert_1$ The servant $\vert_2$ cleaned $\vert_3$ the table $\vert_4$ *but* $\vert_5$ the floor $\vert_6$ **was** $\vert_7$ cleaned $\vert_8$ by $\vert_9$ the maid. $\vert$

"|" indicates segmentation during the presentation; subscripted numbers indicate word position. The underscored word indicates the start of the ambiguity in the a-conditions, bold indicates the disambiguating critical regions. These markings are for purpose of illustration only and did not appear in the actual materials.

Table 2. Results from the linear mixed effects model for critical word position, *wh*-conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	0.045	0.013	3.331	0.001**
Condition	0.045	0.015	2.910	0.005**
Number of *What* seen	-0.033	0.012	-2.810	0.008**
Language Group	-0.039	0.023	-1.652	0.103
Condition × Nr. of *What* seen	-0.024	0.020	-1.221	0.227
Condition × Language Group	0.014	0.027	0.531	0.597
Nr. of *What* seen × Language Group	-0.012	0.020	-0.628	0.532
Condition × Nr. of *What* seen × Language Group	0.000	0.036	0.007	0.995

Note: Nr. of *What* seen: number of preceding *what* sentences seen, log transformed; Condition: *Whether vs. What*; Language Group: *Native English* vs. *L2*. Model: LogRTresidual ~ Condition * Nr. of *What* seen * Language Group + (1 + Condition * Nr. of *What* seen | Subject) + (1 + Condition * Nr. of *What* seen | Item); 79 subjects; 36 items; Log-Likelihood: -691.0. For this model and the models presented in following tables, *P*-values were determined by LmerTest. ** $p < 0.01$, * $p < 0.05$, $^{+}p < 0.1$.

Table 3. Results from the linear mixed effects model for the native English group, *wh*-conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	0.064	0.013	5.031	0.000**
Condition	0.039	0.021	1.872	0.069^{+}
Number of *What* seen	-0.028	0.014	-1.924	0.064^{+}
Condition × Nr. of *What* seen	-0.027	0.027	-0.977	0.332

Note: Nr. of *What* seen: number of preceding *what* sentences seen, log transformed; Condition: *Whether vs. What.* Model: LogRTresidual ~ Condition * Nr. of *What* seen + (1 + Condition * Nr. of *What* seen | Subject) + (1 + Condition * Nr. of *What* seen | Item); 40 subjects; 36 Items; Log-Likelihood: -206.4.

Table 4. Results from the linear mixed effects model for the L2 group, *wh*-conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	0.024	0.023	1.052	0.298
Condition	0.056	0.019	2.960	0.008**
Number of *What* seen	-0.036	0.015	-2.468	0.022*
LexTale	0.001	0.002	0.711	0.482
Condition × Nr. of *What* seen	-0.026	0.026	-1.003	0.316
Condition × LexTale	0.000	0.002	-0.171	0.866
Nr. of *What* seen × LexTale	0.002	0.001	1.660	0.105
Condition × Nr. of *What* seen × LexTale	-0.001	0.002	-0.609	0.543

Note: Nr. of *What* seen: number of preceding *what* sentences seen, log transformed; Condition: *Whether vs. What.* Model: LogRTresidual ~ Condition * Nr. of *What* seen * LexTale + (1 + Condition * Nr. of *What* seen || Subject) + (1 + Condition * Nr. of *What* seen || Item); 39 subjects; 36 Items; Log-Likelihood -459.8

Table 5. Results from the linear mixed effects model, coordination conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	0.008	0.009	0.941	0.351
Condition	0.024	0.010	2.410	0.018*
Number of *And* seen	0.002	0.007	0.215	0.830
Language Group	-0.044	0.014	-3.211	0.002**
Condition × Nr. of *And* seen	-0.011	0.014	-0.805	0.421
Condition × Language Group	0.005	0.020	0.253	0.801
Nr. of *And* seen × Language Group	0.008	0.014	0.553	0.581
Condition × Nr. of *And* seen × Language Group	0.070	0.027	2.592	0.010*

Note: Nr. of *And* seen: number of preceding *And* sentences seen, log transformed. Condition: *But vs. And*; Language Group: *Native English* vs. *L2*. Model: LogRTresidual ~ Condition * Nr. of *And* seen * Language Group + (1 + Condition * Nr. of *And* seen | Subject) + (1 + Condition * Nr. of *And* seen | Item); 79 subjects; 36 items; Log-Likelihood: -107.0.

Table 6. Results from the linear mixed effects model for the native English group, coordination conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	0.030	0.011	2.764	0.008**
Condition	0.021	0.013	1.568	0.125
Number of *And* seen	-0.003	0.009	-0.358	0.720
Condition × Nr. of *And* seen	-0.046	0.017	-2.623	0.019*

Note: Nr. of *And* seen: number of preceding *And* sentences seen, log transformed. Condition: *But vs. And*. Model: LogRTresidual ~ Condition * Nr. of *And* seen + (1 + Condition * Nr. of *And* seen || Subject) + (1 + Condition * Nr. of *And* seen || Item); 40 subjects; 36 Items; Log-Likelihood: 77.1.

Table 7. Results from the linear mixed effects model for the L2 group, coordination conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	-0.015	0.011	-1.373	0.178
Condition	0.029	0.015	1.991	0.054[+]
Number of *And* seen	0.004	0.011	0.387	0.701
LexTale	0.001	0.001	1.117	0.271
Condition × Nr. of *And* seen	0.025	0.021	1.192	0.234
Condition × LexTale	-0.001	0.001	-0.513	0.611
Nr. of *And* seen × LexTale	-0.001	0.001	-1.071	0.292
Condition × Nr. of *And* seen × LexTale	0.001	0.002	0.281	0.778

Note: Nr. of *And* seen: number of preceding *And* sentences seen, log transformed. Condition: *But vs. And*. Model: LogRTresidual ~ Condition * Nr. of *And* seen * LexTale+ (1 + Condition * Nr. of *And* seen || Subject) + (1 + Condition * Nr. of *And* seen || Item); 39 subjects; 36 Items; Log-Likelihood: -168.4

Table 8. Mean accuracy (SD) on the comprehension questions across the experimental conditions

	What	*Whether*	*And*	*But*
Native group	0.90 (0.10)	0.82 (0.14)	0.88 (0.11)	0.89 (0.09)
L2 group	0.84 (0.12)	0.79 (0.15)	0.84 (0.11)	0.85 (0.11)

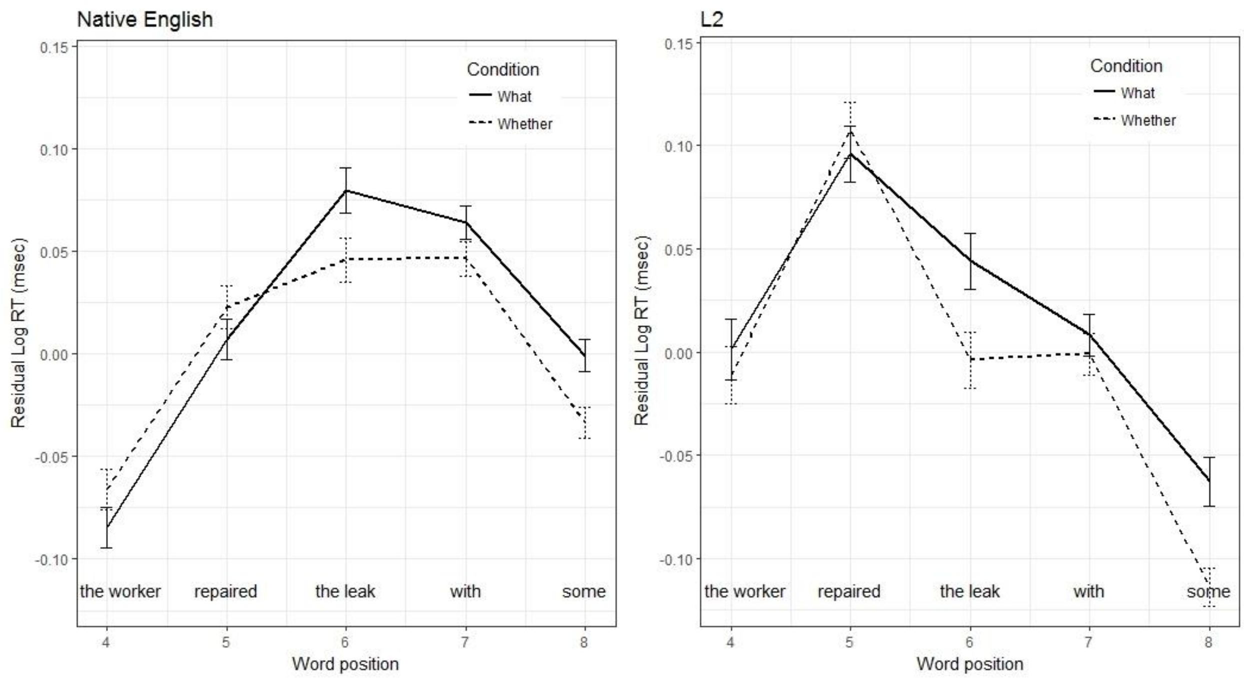

Figure 1: Mean log residual reading times for the *what* and *whether* sentences. Word position 6 corresponds to the critical noun phrase, see Table 1. Left panel: Native English speakers; Right panel: L2 group.

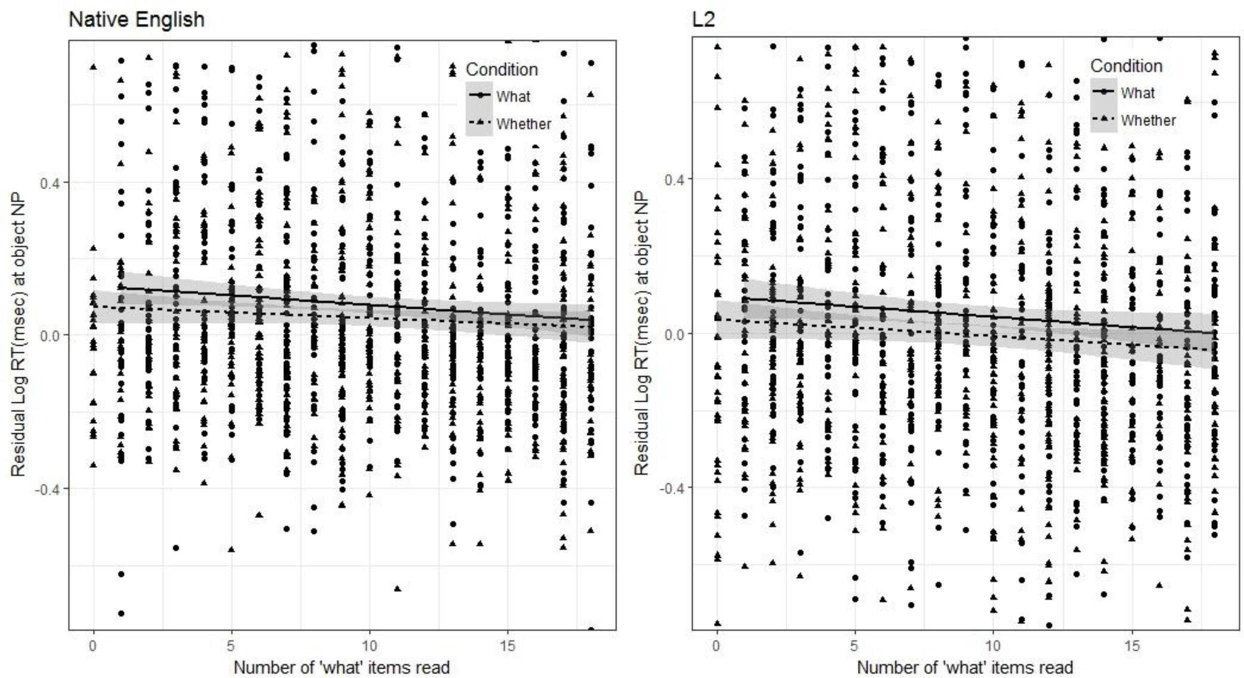

Figure 2 : Mean log residual reading times for the *what* and *whether* conditions at the critical noun phrase as a function of the number of *what* sentences seen in the study. Left panel: Native English speakers; Right panel: L2 group.

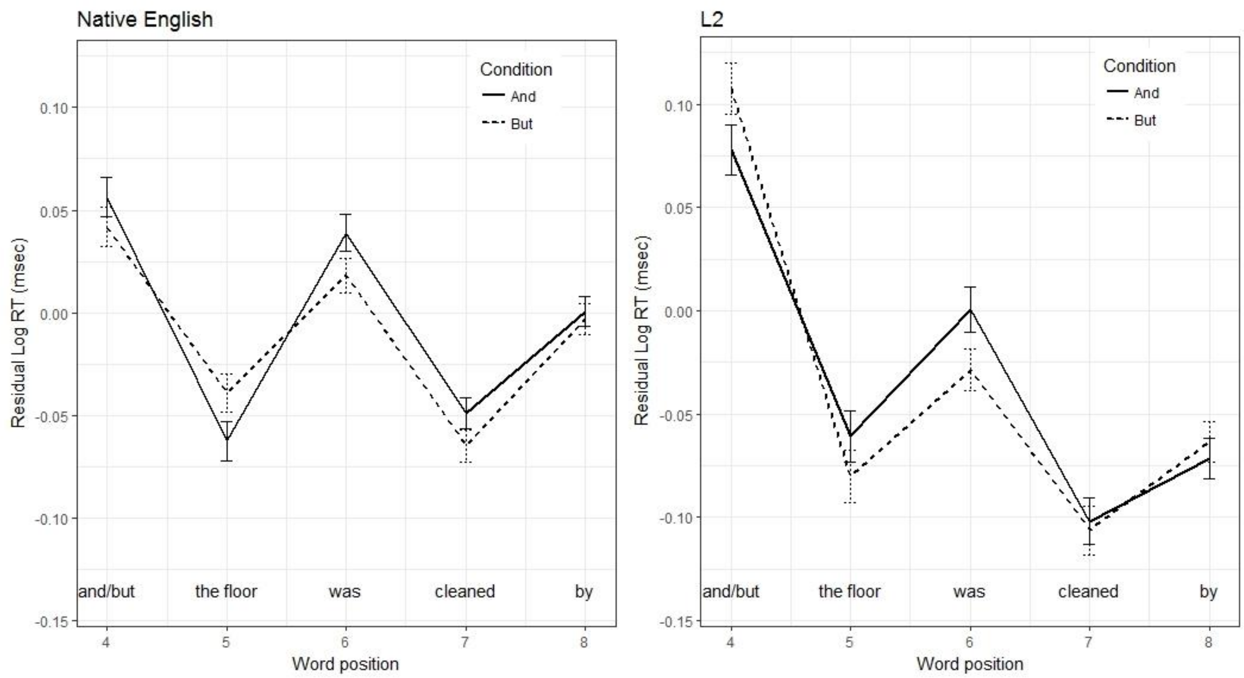

Figure 3: Mean log residual reading times for the *and* and *but* sentences. Word position 6 corresponds to the critical verb, see Table 1. Left panel: Native English speakers; Right panel: L2 group.

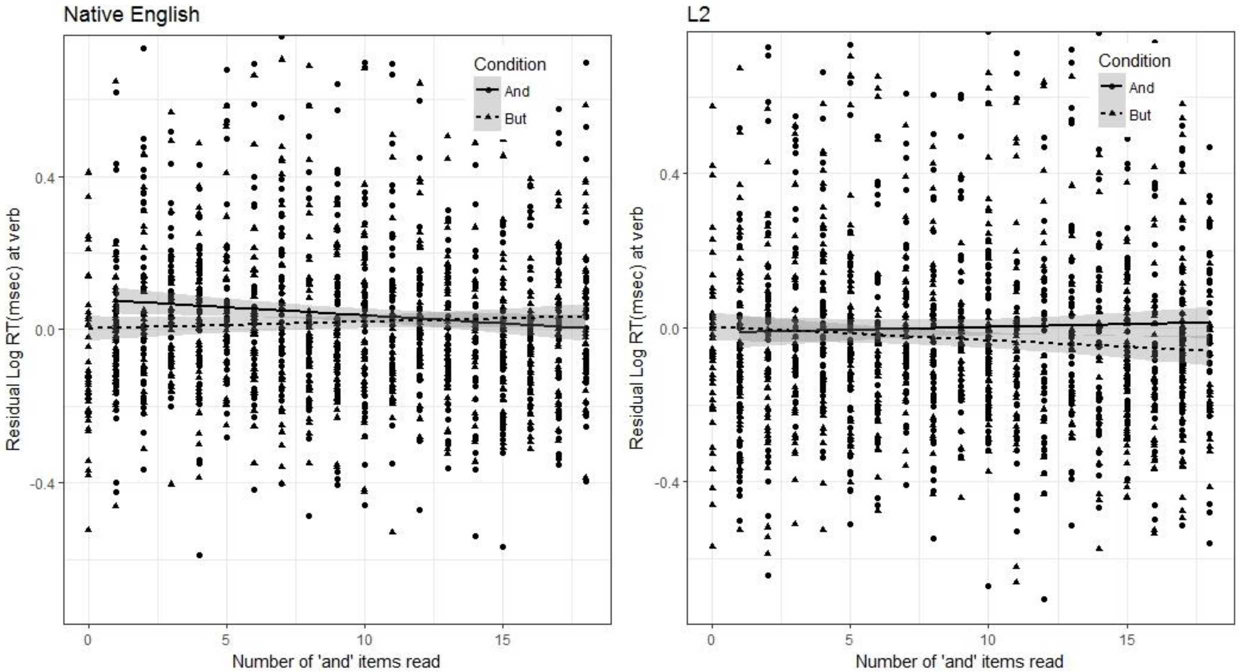

Figure 4: Mean log residual reading times for the *and* and *but* condition at the critical verb as a function of the number of *and* sentences seen in the study. Left panel: Native English speakers; Right panel: L2 group.

Supplementary materials

Supplementary materials A: Experimental materials

Sentences and comprehension questions used in the study. "_" indicated which words were presented in one frame.

Wh-conditions

"what" version is in a; "whether" version is in b.

1a. The_nurse asked what the_doctor prescribed the_medicine for after the_patient_left.

Did the nurse ask about the prescription? Y

1b. The_nurse asked whether the_doctor prescribed the_medicine for the flu after the_patient_left.

Did the nurse ask about the prescription? Y

2a. The_student wondered what the_professor created the_assignment for at the_end_of_the_year.

2b. The_student wondered whether the_professor created the_assignment for the course at the_end_of_the_year.

3a. The_writer wanted_to_know what the_publisher approved the_article for after the_meeting.

Did the writer approve the article? N

3b.The_writer wanted_to_know whether the_publisher approved the_article for the journal after the_meeting.

Did the writer approve the article? N

4a. The_reporter asked what the_player caught the_ball with to block_the_shot.

Did the player drop the ball? N

4b. The_reporter asked whether the_player caught the_ball with his hands to block_the_shot.

Did the player drop the ball? N

5a. The_guard asked what the_thief broke the_glass with to get_into_the_museum.

Did the thief break the glass? Y

5b. The_guard asked whether the_thief broke the_glass with a brick to get_into_the_museum.

Did the thief break the glass? Y

6a. The_guide wondered what the_tourist recognized the_street from during the_bus_tour.

Did the tourist take a bus tour? Y

6b. The_guide wondered whether the_tourist recognized the_street from the magazine during the_bus_tour.

Did the tourist take a bus tour? Y

7a. The_builder wondered what the_worker repaired the_leak with before going_home.

Did the worker repair the leak? Y

7b. The_builder wondered whether the_worker repaired the_leak with some tape before going_home.

Did the worker repair the leak? Y

8a. The_biker wanted_to_know what the_worker blocked the_road for during the_day.

8b. The_biker wanted_to_know whether the_worker blocked the_road for the construction during the_day.

9a. The_supervisor wondered what the_employee reached the_shelf with to get_to_some_boxes.

Did the supervisor reach the shelf? N

9b. The_supervisor wondered whether the_employee reached the_shelf with a ladder to get_to_some_boxes.

Did the supervisor reach the shelf? N

10a. The_operator asked what the_caller requested the_ambulance for on the_college_campus.

10b. The_operator asked whether the_caller requested the_ambulance for the accident on the_college_campus.

11a. The_employer asked what the_company produced the_parts for because nobody_could_say.

11b. The_employer asked whether the_company produced the_parts for the machine because nobody_could_say.

12a. The_man wondered what the_firefighter saved the_documents from when the_alarm_sounded.

Did the firefighter save the man? N

12b. The_man wondered whether the_firefighter saved the_documents from the fire when the_alarm_sounded.

Did the firefighter save the man? N

13a. The_colleague wanted_to_know what the_scientist invented the_procedure for after the_presentation.

Did the colleague invent the procedure? N

13b.The_colleague wanted_to_know whether the_scientist invented the_procedure for the analysis after the_presentation.

Did the colleague invent the procedure? N

14a. The_director asked what the_worker unlocked the_door with that Monday_morning.

Did the director unlock the door? N

14b. The_director asked whether the_worker unlocked the_door with the key that Monday_morning.

Did the director unlock the door? N

15a. The_parent wanted_to_know what the_child watched the_movie about on TV_last_night.

Did the child watch a movie? Y

15b. The_parent wanted_to_know what the_child watched the_movie about on TV_last_night.

Did the child watch a movie? Y

16a. The_woman wondered what the_maid washed the_shirt with to get_rid_of_the_stains.

Did the woman wash the rug? N

16b. The_woman wondered whether the_maid washed the_shirt with some soap to get_rid_of_the_stains.

Did the woman wash the rug? N

17a. The_scientist asked what the_assistant analyzed the_data for in the_lab.

17b. The_scientist asked whether the_assistant analyzed the_data for the experiment in the_lab.

18a. The_assistant wondered what the_host cancelled the_event for after the_call.

18b. The_assistant wondered whether the_host cancelled the_event for the company after the_call.

19a. The_manager asked what the_musician carried the_guitar in on the_tour.

Did the musician bring the guitar? Y

19b. The_manager asked whether the_musician carried the_guitar in the case on the_tour.

Did the musician bring the guitar? Y

20a. The_customer wanted_to_know what the_banker checked the_account for after the_call.

20b.The_customer wanted_to_know whether the_banker checked the_account for a problem after the_call.

21a. The_director wondered what the_assistant controlled the_lights with during the_play.

21b. The_director wondered whether the_assistant controlled the_lights with a switch during the_play.

22a. The_neighbor asked what the_boy crossed the_street for in spite_of_the_traffic.

Did the neighbor cross the street? N

22b. The_neighbor asked whether the_boy crossed the_street for his dog in spite_of_the_traffic.

Did the neighbor cross the street? N

23a. The_nanny wondered what the_boy destroyed the_toy with while he_was_playing.

Did the nanny destroy the toy? N

23b. The_nanny wondered whether the_boy destroyed the_toy with his hands while he_was_playing.

Did the nanny destroy the toy? N

24a. The_guest wanted_to_know what the_waiter lit the_candle with for the_diners.

Did the waiter light the candle? Y

24b.The_guest wanted_to_know whether the_waiter lit the_candle with a match for the_diners.

Did the waiter light the candle? Y

25a. The_customer wondered what the_jeweler needed the_metal for to sell_in_the_store.

Did the jeweler need the metal? Y

25b. The_customer wondered whether the_jeweler needed the_metal for the ring to sell_in_the_store.

Did the jeweler need the metal? Y

26a. The_buyer asked what the_artist painted the_wall with at the_event.

26b. The_buyer asked whether the_artist painted the_wall with a brush at the_event.

27a. The_boy wondered what the_teacher solved the_problem with when it_was_discussed_in_class.

Did the boy solve the problem? N

27b. The_boy wondered whether the_teacher solved the_problem with a calculator when it_was_discussed_in_class.

Did the boy solve the problem? N

28a. The_programmer asked what the_teenager updated the_software for on the_computer.

Did the programmer update the software? N

28b. The_programmer asked whether the_teenager updated the_software for the app on the_computer.

Did the programmer update the software? N

29a. The_student wanted_to_know what the_instructor reviewed the_essay for since no_grades_were_posted.

Did the student review the essay? N

29b. The_student wanted_to_know whether the_instructor reviewed the_essay for the contest since no_grades_were_posted.

Did the student review the essay? N

30a. The_designer asked what the_actor wore the_hat to the other_night.

Did the actor wear the hat? Y

30b. The_designer asked whether the_actor wore the_hat to the party the other_night.

Did the actor wear the hat? Y

31a. The_customer wondered what the_worker removed the_battery from at the_store.

Did the worker remove the battery? Y

31b. The_customer wondered whether the_worker removed the_battery from the phone at the_store.

Did the worker remove the battery? Y

32a. The_passenger wanted_to_know what the_pilot delayed the_flight for but nobody_knew.

Did the pilot delay the flight? Y

32b. The_passenger wanted_to_know whether the_pilot delayed the_flight for the luggage but nobody_knew.

Did the pilot delay the flight? Y

33a. The_engineer asked what the_assistant damaged the_computer with late last_night.

33b. The_engineer asked whether the_assistant damaged the_computer with the virus late last_night.

34a. The_stranger wondered what the_beggar accepted the_money for at the_bus_stop.

Did the beggar accept the money? Y

34b. The_stranger wondered whether the_beggar accepted the_money for some food at the_bus_stop.

Did the beggar accept the money? Y

35a. The_researcher wanted_to_know what the_professor attended the_conference for when they_met.

35b. The_researcher wanted_to_know whether the_professor attended the_conference for the_talk when they_met.

36a. The_policeman asked what the_driver examined the_car for after the_crash.

36b. The_policeman asked whether the_driver examined the_car for any scratches after the_crash.

Coordination-conditions

"and" version is in a; "but" version is in b.

101a. The_biologist studied the_plant and its_flower was very hard to see.

101b. The_biologist studied the_plant but its_flower was very hard to see.

102a. The_woman advertised the_concert and the_festival was completely sold_out last week.

102b. The_woman advertised the_concert but the_festival was completely sold_out last week.

103a. The_maid folded the_blanket and the_laundry was put in a_big basket.

Did the maid fold the sheet? N

103b. The_maid folded the_blanket but the_laundry was put in a_big basket.

Did the maid fold the sheet? N

104a. The_magician performed the_trick and the_joke was later used in a_TV_show.

104b. The_magician performed the_trick but the_joke was later used in a_TV_show.

105a. The_woman baked the_cake and the_pie was made by her daughter.

Did the woman bake the pie? N

105b. The_woman baked the_cake but the_pie was made by her daughter.

Did the woman bake the pie? N

106a. The_sports_fans watched the_match and the_commentary was shown on another channel.

106b. The_sports_fans watched the_match but the_commentary was shown on another channel.

107a. The_man purchased the_painting and the_drawing was bought by his wife.

107b. The_man purchased the_painting but the_drawing was bought by his wife.

108a. The_businessman replied_to the_email and the_phone_call was answered by his secretary.

Did the businessman return the call? N

108b. The_businessman replied_to the_email but the_phone_call was answered by his secretary.

Did the businessman return the call? N

109a. The_trainer planned the_workout and the_schedule was made by the coach.

Did the coach determine the schedule? Y

109b. The_trainer planned the_workout but the_schedule was made by the coach.

Did the coach determine the schedule? Y

110a. The_candy_maker filled the_bowl and the_jar was completely filled with chocolate_bars.

110b. The_candy_maker filled the_bowl but the_jar was completely filled with chocolate_bars.

111a. The_boy rolled_up the_carpet and the_rug was moved by the girl.

Did the girl move the rug? Y

111b. The_boy rolled_up the_carpet but the_rug was moved by the girl.

Did the girl move the rug? Y

112a. The_chef stirred the_soup and the_sauce was stirred by the assistant.

Did the chef stir the soup? Y

112b. The_chef stirred the_soup but the_sauce was stirred by the assistant.

Did the chef stir the soup? Y

113a. The_dog buried the_bone and the_stick was left behind the doghouse.

Did the dog bury the stick? N

113b. The_dog buried the_bone but the_stick was left behind the doghouse.

Did the dog bury the stick? N

114a. The_driver delivered the_package and the_letter was brought by the mailman.

Did the driver deliver the package? Y

114b. The_driver delivered the_package but the_letter was brought by the mailman.

Did the driver deliver the package? Y

115a. The_clerk sold the_phone and the_iPad was sold by his colleague.

Did the clerk sell the iPad? N

115b. The_clerk sold the_phone but the_iPad was sold by his colleague.

Did the clerk sell the iPad? N

116a. The_servant cleaned the_table and the_floor was cleaned by the the_maid.

Did the servant clean the floor? N

116b. The_servant cleaned the_table but the_floor was cleaned by the the_maid.

Did the servant clean the floor? N

117a. The_worker organized the_shed and the_garage was emptied to prepare for_renovations.

117b. The_worker organized the_shed but the_garage was emptied to prepare for_renovations.

118a. The_teacher graded the_exam and the_homework was checked by the assistant.

Did the assistant check the homework? Y

118b. The_teacher graded the_exam but the_homework was checked by the assistant.

Did the assistant check the homework? Y

119a. The_girl chased the_cat and the_dog was chasing the little bunny.

Did the girl chase the cat? Y

119b. The_girl chased the_cat but the_dog was chasing the little bunny.

Did the girl chase the cat? Y

120a. The_architect printed the_design and the_map was drawn by the assistant.

Did the assistant draw the map? Y

120b. The_architect printed the_design but the_map was drawn by the assistant.

Did the assistant draw the map? Y

121a. The_zookeeper fed the_giraffe and the_monkey was entertained by the trainer.

Did the zookeeper entertain the monkey? N

121b. The_zookeeper fed the_giraffe but the_monkey was entertained by the trainer.

Did the zookeeper entertain the monkey? N

122a. The_man built the_table and the_bench was carved by the artist.

Did the artist build the table? N

122b. The_man built the_table but the_bench was carved by the artist.

Did the artist build the table? N

123a. The_tutor explained the_question and the_answer was given in the textbook.

Did the tutor explain the question? Y

123b. The_tutor explained the_question but the_answer was given in the textbook.

Did the tutor explain the question? Y

124a. The_pirate hid the_silver and the_gold was kept in the ship.

Did the pirate hide the gold? N

124b. The_pirate hid the_silver but the_gold was kept in the ship.

Did the pirate hide the gold? N

125a. The_professor wrote the_paper and the_book was published by the journalist.

Did the professor publish the book? N

125b. The_professor wrote the_paper but the_book was published by the journalist.

Did the professor publish the book? N

126a. The_boy received the_card and the_gift was delivered a few_days later.

126b. The_boy received the_card but the_gift was delivered a few_days later.

127a. The_waiter served the_meal and the_dessert was presented by the chef.

Did the chef present the meal? N

127b. The_waiter served the_meal but the_dessert was presented by the chef.

Did the chef present the meal? N

128a. The_man read the_note and the_letter was mailed to the office.

128b. The_man read the_note but the_letter was mailed to the office.

129a. The_scientist mixed the_powder and the_liquid was poured in the bucket.

Did the scientist mix the powder? Y

129b. The_scientist mixed the_powder but the_liquid was poured in the bucket.

Did the scientist mix the powder? Y

130a. The_child planted the_flower and the_tree was planted by the gardener.

Did the gardener plant the tree? Y

130b. The_child planted the_flower but the_tree was planted by the gardener.

Did the gardener plant the tree? Y

131a. The_nanny prepared the_snack and the_drink was left in the fridge.

131b. The_nanny prepared the_snack but the_drink was left in the fridge.

132a. The_customer signed the_form and the_contract was signed by the manager.

Did the customer sign the contract? N

132b. The_customer signed the_form but the_contract was signed by the manager.

Did the customer sign the contract? N

133a. The_musician played the_piano and the_violin was tuned by the director.

Did the director tune the violin? Y

133b. The_musician played the_piano but the_violin was tuned by the director.

Did the director tune the violin? Y

134a. The_engineer sent the_code and the_password was changed in the system.

Did the engineer send the code? Y

134b. The_engineer sent the_code but the_password was changed in the system.

Did the engineer send the code? Y

135a. The_seller advertised the_product and the_discount was applied at the register.

135b. The_seller advertised the_product but the_discount was applied at the register.

136a. The_officer stamped the_passport and the_form was inspected by a_second officer.

136b. The_officer stamped the_passport but the_form was inspected by a_second officer.

Distractor Items

301. The_tourist wondered when the_train would arrive from Paris.

302. The_servant asked when the_suite in the_hotel might be ready so_the_guest_could_check_in.

303. The_advisor wanted_to_know when the_grade from the_exam could be sent to_the_administrators.

304. The_librarian told_us when the_book from the_collection might be returned.

305. The_child wondered when the_cookie from the_jar was eaten.

306. The_student asked when the_meeting for the_group would be held that_day.

307. The_lifeguard wondered why the_pool in the_neighborhood was closed since_nobody_knew.

Was the pool open? N

308. The_artist asked_ why the_supplies were missing from the studio.

Were the supplies in the studio? N

309. The_guide wanted_to_know why the_tour for the_castle would be delayed that_afternoon.

Was the tour on time? N

310. The_photographer told_us why the_photograph in the_album was ripped.

Was the photograph ripped? Y

311. The_customer wanted_to_know why the_cost for the_trip would increase after the break.

312. The_driver told_us why the_bus for the_students was late at_the_bus_stop.

313. The_man wondered how the_decision for the_trial would be made given_the_lack_of_evidence.

314. The_mathematician asked how the_solution in the_manual was described since_he_was_curious.

315. The_guard wanted_to_know how the_alarm in the_museum could be turned off.

316. The_biker told_us_ how the_trail was designed to avoid traffic.

317. The_policeman wondered how the_car in the_garage was stolen last night.

Was the car stolen? Y

318. The_manager asked how the_necklace in the_store disappeared from the_table.

319. The_nanny told_us that the_cake had been in the_oven for_an_hour but_it_was_still_raw.

320. The_assistant mentioned that the_mouse had escaped from the_cage this_morning.

Did the mouse escape from the cage? Y

321. The_director said that the_actress apologized for the_mistake during the_play.

322. The_businessman saw that the_meeting had ended with an_agreement.

323. The_doctor told_us that the_disease had gotten worse over the_years.

Did the disease get worse? Y

324. The_seller mentioned that the_statue sold for thousands of dollars.

325. The_journalist said that the_event took place while it was raining.

Was it raining during the event? Y

326. The_pilot saw that the_weather was too stormy for the_plane to take_off.

327. The_farmer told_us that the_cow had been frightened by the_lightning last_night.

328. The_nanny mentioned that the_show had been playing all day on television.

329. The_buyer said that the_sign fell apart by the_stand next to the_snacks.

Did the sign fall apart? Y

330. The_staff saw that the_supply decreased since the_last time the_store opened.

331. The_programmer told_us that the_network crashed after the_storm last night.

Was the network working after the storm? N

332. The_woman mentioned that the_artwork was broken after someone knocked it over.

333. The_reporter said that the_sun rose earlier than usual this morning.

334. The_gardener saw that the_tree had grown a_lot since last_month.

335. The_mother told_us that the_boy was bitten by a_spider.

Did a bee sting the boy? N

336. The_friend mentioned that the_girl fell off the_bike while playing outside.

Did the girl fall off the scooter? N

401. The_deer ran into the_forest after the_hunter tried to shoot it.

402. The_baby played on the_blanket while the_grandmother sipped her coffee.

403. The_worker knocked on the_door before the_director invited him in.

Did the worker knock on the door? Y

404. The_speaker explained the_idea as the_people listened to the_speech.

405. The_stylist complained about the_price while the_assistant nodded her head in_agreement.

Did the assistant complain? N

406. The_client sat in the_chair as the_man cut his hair.

407. The_athlete trained in the_gym while the_trainer planned his meals.

Did the athlete plan his meals? N

408. The_dentist looked over the_x-ray as the_patient worried about the_cost.

409. The_mother laughed at the_story while the_child looked unhappy.

410. The_surgeon appeared at the_hospital after the_nurse arranged the_equipment for the_surgery.

411. The_engineer gazed at the_plan after the_architect explained the_notes on the_layout.

Did the architect explain the plan? N

412. The_man had just reached the_shelter as the_wind blew the_roof off the_building.

413. The_girl was_worried about her_grades while her_friend ignored the_test_scores.

Did the friend ignore the scores? Y

414. The_student graduated from the_university as his_parents attended the_ceremony.

415. The_daughter helped with the_dinner while the_mother prepared the_table for the_family.

Did the daughter prepare the table? N

416. The_boss paid for the_coffee while the_employee checked his calendar.

417. The_city recovered from the_earthquake as the_workers fixed the_roofs of the_buildings.

418. The_banker responded to the_alert while the_bank prevented the_robbery.

Was the bank robbed? N

419. As the_rabbit was eating the_carrot the_fox came closer.

Did the rabbit come closer? N

420. While the_family was enjoying the_dinner the_pet stared at its_bowl.

421. While the_boy practiced the_song the_relatives listened with excitement.

Were the relatives excited? Y

422. As the_parent started packing the_clothes the_children offered to help.

423. After the_students had handed_in the_test the_teacher explained the answers.

424. Even_though the_man was using a_map he managed to get lost.

425. When the_family planned the_wedding a_friend agreed to take pictures.

426. While the_child was holding the_fork the_parents watched with excitement.

427. As the_celebrity was walking the_streets his_guards formed a_circle around him.

428. While the_cat was climbing the_tree the_kids shouted for its attention.

429. As the_diver approached the_water the_crowd watched in amazement.

Was anyone watching the diver? Y

430. While the_man was testing the_machine his_friend made a_suggestion.

431. As their_son studied abroad the_parents worried about his_safety.

432. Although the_girl wanted a_pony her_parents ignored her wish.

Did the girl wish for a pony? Y

433. Even_though the_actor forgot his_lines the_director hired him for the_play.

434. Although the_boy took the_exam the_teacher could not pass him.

Did the boy fail? Y

435. While the_worker was opening the_window the_man looked confused.

436. While the_man was visiting the_city his_friends suggested a restaurant for lunch.

Supplementary materials B: Analysis using non-transformed raw reading times

The analysis reported in the main text uses log-transformed residual reading times. Below we report analyses using non-transformed raw reading times. We used the same cutoff procedure as in the main text, in which we first omitted all data points that were shorter than 100 ms or longer than 5000, and then omitted data points exceeding the mean plus 2.5 standard deviations for each individual. The main differences with the analysis reported in the main text are that, first, in the analysis on the raw reading times, an adaptation effect was seen in the *wh*-conditions (interaction of Condition by Number of *What* items seen); this effect was also present in the by-group analysis of the native English group, but not in the L2 group. Second, in the coordination conditions the triple interaction of Condition by Language Group by Number of *And* items was weaker. See footnote 2 in the main text for discussion.

Figure SB.1

Raw reading times for selected word positions for the native English and L2 group, *wh* conditions. The critical position is position 6. Note that the y-axis starts at different points for the two groups.

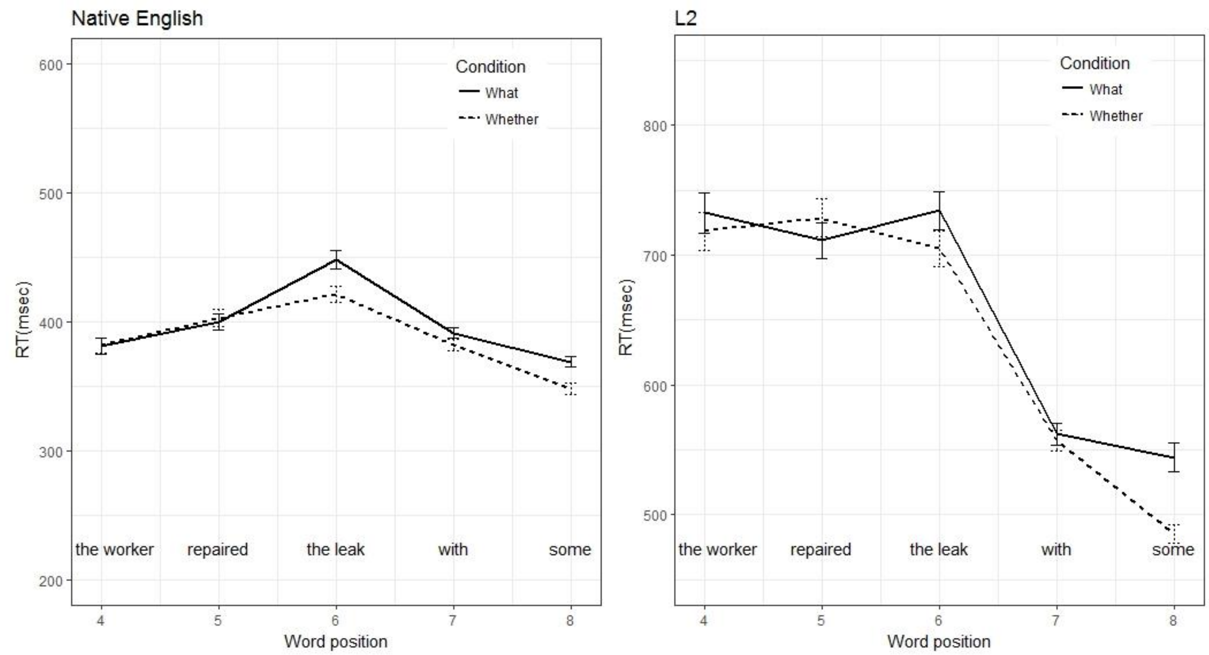

Figure SB.2

Raw reading times for position 6 in the *wh*-condition as a function of the number of *what* sentences seen, for the native English and L2 groups.

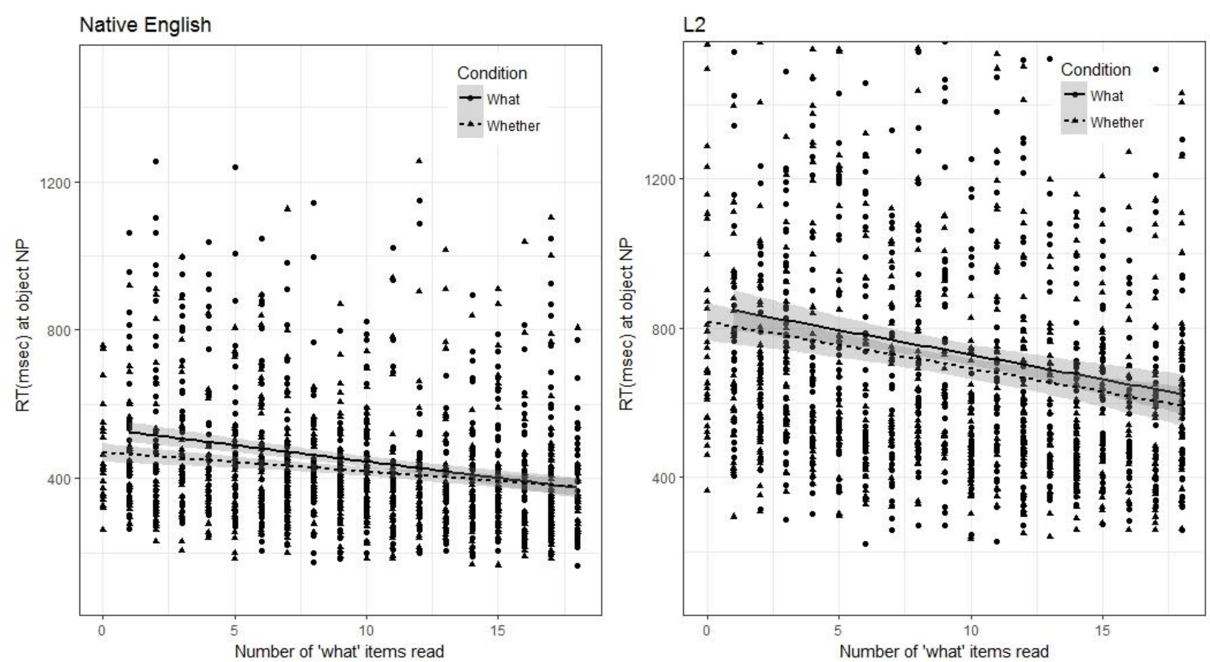

Figure SB.3

Raw readings time for selected word positions for the native English and L2 group, coordination-conditions. The critical position is position 6. Note that the y-axis starts at different points for the two groups.

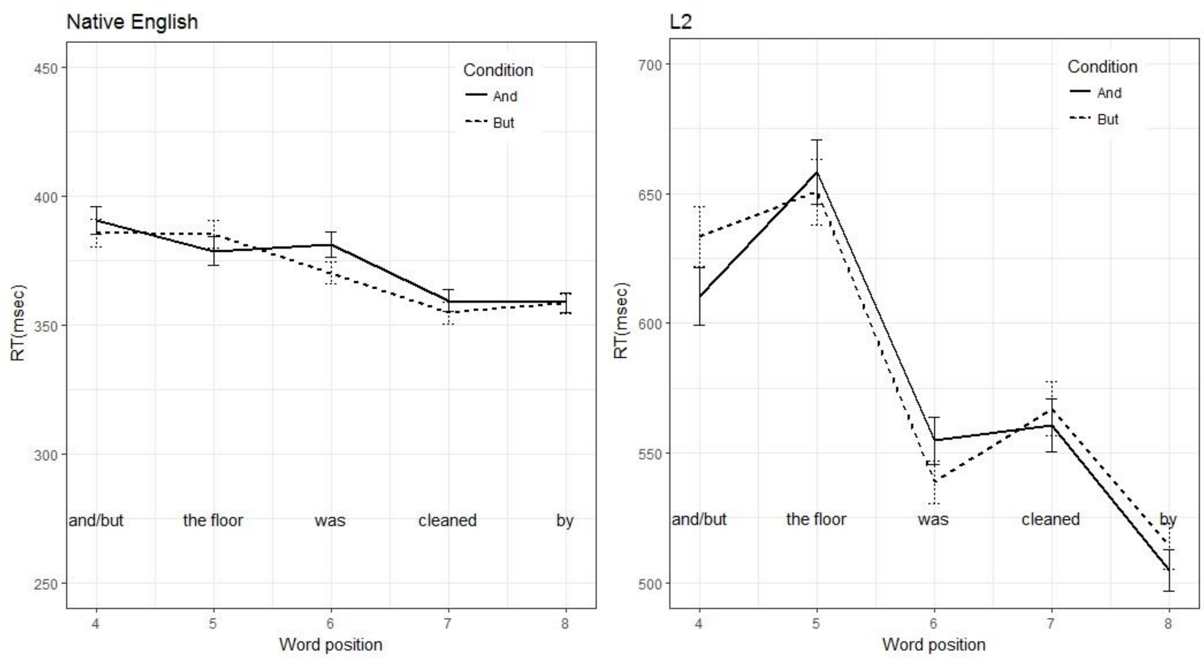

Figure SB.4

Raw reading times for position 6 in the coordination-conditions as a function of the number of *and*-sentences seen, for the native English and L2 groups. Note that the y-axis starts at different points for the two groups.

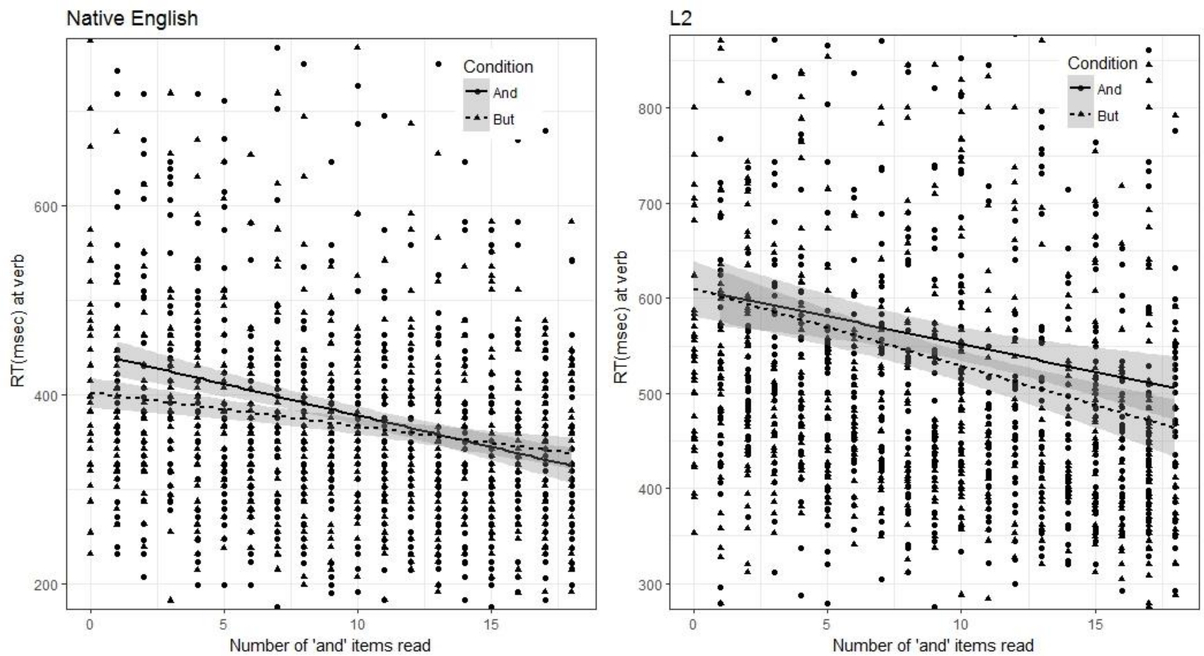

Table S1. Results from the linear mixed effects model using raw reading times for critical word position, *wh*-conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	580.389	20.379	28.480	0.000***
Condition	39.494	10.475	3.770	0.000***
Number of *What* seen	-86.261	11.275	-7.651	0.000***
Language Group	285.805	39.547	7.227	0.000***
Condition × Nr. of *What* seen	-32.552	15.773	-2.064	0.043*
Condition × Language Group	21.586	20.479	1.054	0.294

Nr. of *What* seen × Language Group	-42.417	20.080	-2.112	0.038*
Condition × Nr. of *What* seen × Language Group	-7.018	30.312	-0.232	0.817

Note: Nr. of *What* seen: number of preceding *what* sentences, log transformed. Model: RT ~ Condition * Nr. of *What* seen * Language Group + (1 + Condition * Nr. of *What* seen | Subject) + (1 + Condition * Nr. of *What* seen | Item); 79 subjects; 36 items ; Log-Likelihood: -18804.8. In this and other tables, *P*-values were determined by lmerTest. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, $^{+}p < 0.1$.

Table S2. Results from the linear mixed effects model using raw reading times for critical word position, for the native English group, *wh*-conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	435.881	16.385	26.603	0.000***
Condition	28.949	9.037	3.203	0.003**
Number of *What* seen	-62.861	7.725	-8.138	0.000***
Condition × Nr. of *What* seen	-32.283	12.128	-2.662	0.008**

Model: RT ~ Condition * Nr. Of *What* seen + (1 + Condition * Nr. Of *What* seen || Subject) + (1 + Condition * Nr. Of *What* seen || Item); 40 subjects; 36 Items; Log-Likelihood: -8828.3.

Table S3. Results from the linear mixed effects model using raw reading times for critical word position, for the L2 group, *wh*-conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	726.259	35.634	20.381	0.000***

Condition	52.067	18.269	2.850	0.007**
Number of *What* seen	-105.910	18.563	-5.705	0.000***
LexTale	-5.790	3.099	-1.869	0.069+
Condition × Nr. of *What* seen	-36.713	26.886	-1.366	0.180
Condition × LexTale	-1.143	1.600	-0.714	0.479
Nr. of *What* seen× LexTale	2.991	1.535	1.949	0.059+
Condition × Nr. of *What* seen × LexTale	0.495	2.363	0.210	0.835

Model: RT ~ Condition * Nr. Of *What* seen + (1 + Condition * Nr. Of *What* seen | Subject) + (1 + Condition * Nr. Of *What* seen | Item); 39 subjects; 36 Items; Log-Likelihood: -9678.6.

Table S4. Results from the linear mixed effects model using raw reading times for critical word position, coordination-conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	460.751	10.815	42.604	0.000***
Condition	21.338	6.273	3.402	0.001***
Number of *And* seen	-47.015	6.160	-7.632	0.000***
Language Group	168.180	21.051	7.989	0.000***
Condition × Nr. of *And* seen	-11.524	9.609	-1.199	0.233
Condition × Language Group	9.400	12.537	0.750	0.455
Nr. of *And* seen × Language Group	-15.770	12.302	-1.282	0.204
Condition × Nr. of *And* seen ×	35.377	18.995	1.862	0.065+

Language Group

Note: Nr. of *And* seen: number of preceding *And* sentences, log transformed. Model: RT ~ Condition * Nr. of *And* seen * Language Group + (1 + Condition * Nr. of *And* seen | Subject) + (1 + Condition * Nr. of *And* seen |Item); 79 subjects; 36 items ; Log-Likelihood: -18330.8

Table S5. Results from the linear mixed effects model using raw reading times for critical word position, for the native English group, coordination-conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	376.741	11.261	33.454	0.000***
Condition	15.725	6.082	2.585	0.014*
Number of *And* seen	-38.891	5.948	-6.539	0.000***
Condition × Nr. of *And* seen	-28.282	10.397	-2.720	0.010*

Model: RT ~ Condition * Nr. Of *And* seen + (1 + Condition * Nr. Of *And* seen | Subject) + (1 + Condition * Nr. Of *And* seen | Item); 40 subjects; 36 Items; Log-Likelihood: -8655.9.

Table S6. Results from the linear mixed effects model using raw reading times for critical word position for the L2 group, coordination-conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	545.984	17.754	30.752	0.000***
Condition	26.785	10.551	2.538	0.011*
Number of *And* seen	-55.216	10.908	-5.062	0.000***
LexTale	-2.566	1.553	-1.652	0.107
Condition × Nr. of *And* seen	5.603	15.135	0.370	0.711

Condition × LexTale	-0.508	0.932	-0.545	0.586
Nr. of *And* seen× LexTale	-1.065	0.958	-1.112	0.274
Condition × Nr. of *And* seen × LexTale	0.208	1.320	0.157	0.875

Model: RT ~ Condition * Nr. of *And* seen + (1 + Condition * Nr. of *And* seen || Subject) + (1 + Condition * Nr. of *And* seen || Item); 39 subjects; 36 Items; Log-Likelihood: -9392.3.

Supplementary materials C: Analysis on log residual reading times of the *wh*-conditions, using a more liberal cutoff

A commonly used criterion is to omit data points shorter than 100 ms and longer than 2000 ms, without further cutoff procedures, assuming that the log transformation reduces outliers (Fine et al., 2013). In the analysis reported here, we applied a similar procedure to our data. However, the upper limit of 2000 ms may not be appropriate for L2 learners. We therefore approximated the high cutoff value of 2000 ms by taking the mean plus 6.5 standard deviations for the English speakers (calculated over all data points). This amounted to 1947 ms. We then calculated the mean and standard deviation for the L2 data, and also used the mean plus 6.5 standard deviations as a high cutoff (3880 ms) for the L2 data. In both groups, we also omitted data points shorter than 100 ms. This procedure affected less than 0.5% of the all data points in either group. Restricted to the critical word positions (point of disambiguation), the cutoff also affected less than 0.5% of the data for either group. As in the analysis reported in the main text, we then log-transformed the reading times (natural logarithm) to adjust for the skewedness of the distribution. The Box-Cox procedure (Box & Cox, 1964) confirmed that a log transformation was appropriate (maximal λ was -0.6). For the native and L2 groups separately, we calculated residual reading times based on a linear mixed effects model on all data (experimental items as well as distractors), with the length of the word in the number of characters, and the (natural) log-transformed position of the trial in the experiment as fixed effects. Random effects included by-participant intercepts and by-participant slopes for word length and the log of the trial position. As addressed in the discussion of the main text, the native English group showed adaptation effects in this analysis, but not in the analysis reported in the main text using a more conservative cutoff procedure. This suggests that the adaptation effect is driven by only 5% of the data.

Figure SC.1.

Log residual reading times for selected word positions for the native English and L2 group, *wh* conditions. The critical position is position 6. Note that the y-axis starts at different points for the two groups.

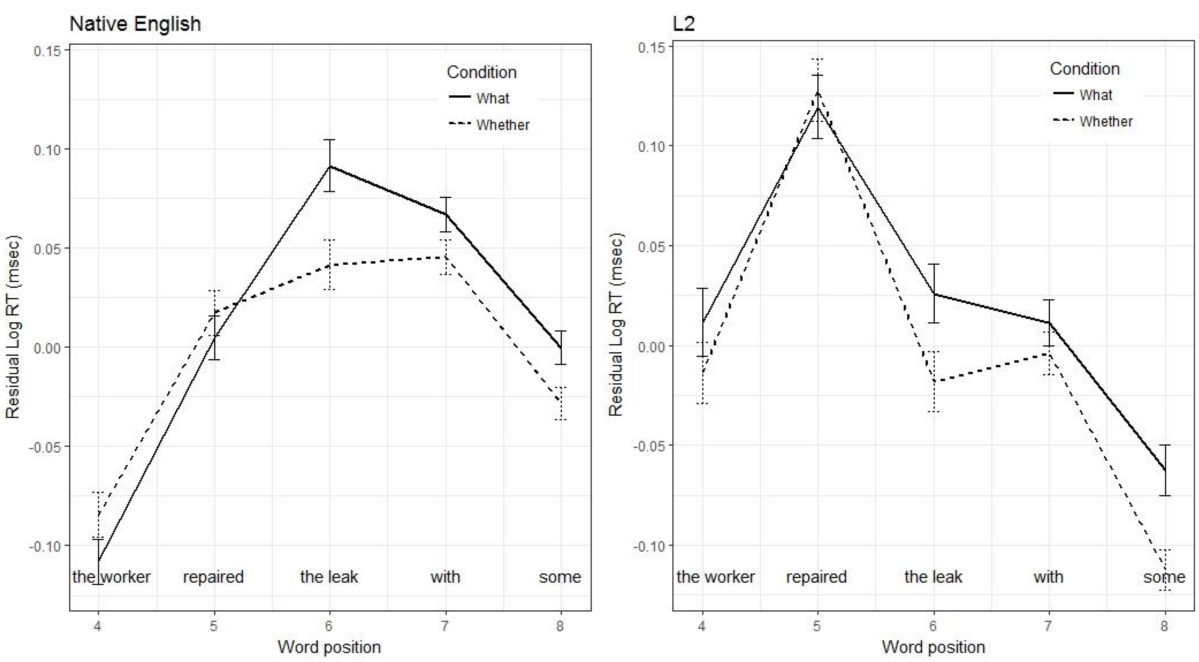

Figure SC.2

Log residual reading times for position 6 in the *wh*-condition as a function of the number of *what* sentences seen, for the native English and L2 groups.

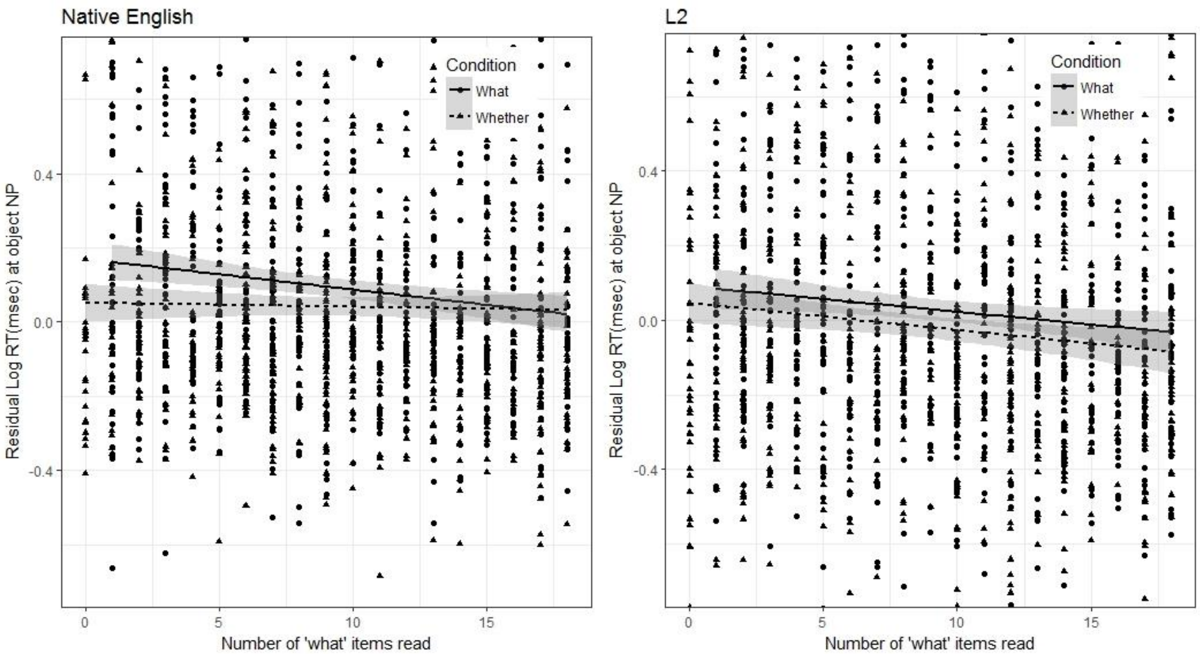

Table S7. Results from the linear mixed effects model using log residual reading times with a liberal cutoff, critical word position, *wh*-conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	0.037	0.016	2.279	0.025*
Condition	0.052	0.016	3.249	0.002**
Number of *What* seen	-0.043	0.014	-3.102	0.003**
Language Group	-0.060	0.027	-2.226	0.029*
Condition × Nr. of *What* seen	-0.047	0.022	-2.163	0.035*
Condition × Language Group	-0.009	0.030	-0.308	0.759
Nr. of *What* seen × Language	-0.022	0.025	-0.871	0.386

Group				
Condition × Nr. of *What* seen × Language Group	0.052	0.039	1.338	0.182

Note: Nr. of *What* seen: number of preceding *what* sentences, log transformed. Model: RT ~ Condition * Nr. of *What* seen * Language Group + (1 + Condition * Nr. of *What* seen | Subject) + (1 + Condition * Nr. of *What* seen | Item); 79 subjects; 36 items ; Log-Likelihood: -1060.1. In this and other tables, *P*-values were determined by lmerTest. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, $^{+}p < 0.1$.

Table S8. Results from the linear mixed effects model for the native English group, *wh*-conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	0.067	0.015	4.424	0.000***
Condition	0.054	0.021	2.525	0.016*
Number of *What* seen	-0.036	0.018	-1.956	0.059+
Condition × Nr. of *What* seen	-0.067	0.032	-2.074	0.043*

Model: RT ~ Condition * Nr. Of *What* seen + (1 + Condition * Nr. Of *What* seen | Subject) + (1 + Condition * Nr. Of *What* seen | Item); 40 subjects; 36 Items; Log-Likelihood: -453.7.

Table S9. Results from the linear mixed effects model for the L2 group, *wh*-conditions

	Estimate	SE	T-value	*p*-value
(Intercept)	0.006	0.026	0.228	0.820
Condition	0.051	0.023	2.282	0.027*
Number of *What* seen	-0.052	0.019	-2.797	0.009**
LexTale	0.001	0.002	0.504	0.617
Condition × Nr. of *What* seen	-0.023	0.029	-0.790	0.431

Condition × LexTale	-0.001	0.002	-0.451	0.654
Nr. of *What* seen× LexTale	0.002	0.002	1.448	0.156
Condition × Nr. of *What* seen × LexTale	-0.002	0.003	-0.680	0.497

Model: RT ~ Condition * Nr. Of *What* seen + (1 + Condition * Nr. Of *What* seen | Subject) + (1 + Condition * Nr. Of *What* seen | Item); 39 subjects; 36 Items; Log-Likelihood: -593.6.

References

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B (Methodological),* 26, 211-252.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE,* 8, 1-18. doi:10.1371/journal.pone.0077661