Maintenance cost in the processing of subject-verb dependencies

Bojana Ristic[1,2], Simona Mancini[1], Nicola Molinaro[1,3] & Adrian Staub[4]

[1] BCBL – Basque Center on Cognition, Brain and Language, University of the Basque Country UPV/EHU

[2] Division of Psychology and Language Sciences, University College London

[3] Ikerbasque, Bilbao, Spain

[4] Department of Psychological and Brain Sciences, University of Massachusetts Amherst, USA

Author note

Data and scripts used in this article are available at: https://doi.org/10.17605/OSF.IO/VS6C5

Correspondence should be addressed to Bojana Ristic, Chandler House, 2 Wakefield Street, London WC1N 1PF; b.ristic@ucl.ac.uk

Abstract

Although research in sentence comprehension has suggested that processing long-distance dependencies involves maintenance between the elements that form the dependency, studies on maintenance of long-distance subject–verb (SV) dependencies are scarce. The few relevant studies have delivered mixed results using self-paced reading or phoneme-monitoring tasks. In the current study, we used eye tracking during reading to test whether maintaining a long-distance SV dependency results in a processing cost on an intervening adverbial clause. In Experiment 1, we studied this question in Spanish and found that both go-past reading times and regressions out of an adverbial clause to the previous regions were significantly increased when the clause interrupts a SV dependency compared to when the same clause doesn't interrupt this dependency. We then replicated these findings in English (Experiment 2), observing significantly increased go-past reading times on a clause interrupting a SV dependency. The current study provides the first eye-tracking data showing a maintenance cost in the processing of SV dependencies cross-linguistically. Sentence comprehension models should account for the maintenance cost generated by SV dependency processing, and future research should focus on the nature of the maintained representation.

Keywords

maintenance cost; subject-verb dependencies; long-distance dependencies; eye-tracking during reading

Research in sentence comprehension has suggested that processing long-distance dependencies involves some sort of maintenance, expectation, or prediction between the elements that form the dependency. However, while research on the processing of filler-gap structures has provided evidence that maintaining a filler-gap dependency results in a processing cost on the intervening elements (e.g. Fiebach, Schlesewsky, & Friederici, 2002; Phillips, Kazanina, & Abada, 2005), studies on the maintenance of long-distance subject-verb dependencies are scarce. The few relevant studies have delivered mixed results, and are open to various interpretations (Chen et al., 2005; Hakes et al., 1976; Van Dyke & Lewis, 2003). In the current study, we use eye-tracking during reading to test whether maintaining a long-distance subject-verb dependency results in a processing cost on the interpolated elements in Spanish and English.

<u>Maintaining subject-verb dependencies: the existing evidence</u>

In a phoneme monitoring task, Hakes et al. (1976) compared sentences containing subject-modifying relative clauses, embedded between the matrix subject and verb (1a) with sentences containing object-modifying clauses, located after all the arguments of the matrix clause (1b).

1a) After the final curtain on opening night, the director *(that) the repertory company had hired* praised the star performer.

1b) After the final curtain on opening night, the star performer praised *the director (that) the repertory company had hired*.

No difference in phoneme monitoring performance was found on the relative clause region. The authors concluded that their results do not support the claim that center-embedded sentences are more difficult to process; in the present context, this suggests that there is no additional cost related to a subject-verb dependency being held over the embedded clause.

In a self-paced reading task, Van Dyke & Lewis (2003) compared processing of embedded clauses that either did clearly intervene between subject and verb (2a, c), or did not appear to do so due to a garden path effect (2b, d). In the conditions (2a, c) where *that* was present, the reader would treat *the man* as a clausal subject, and would therefore be maintaining an expectation for a verb while reading the following material, while in the garden path conditions (2b, d), the reader would be expected to treat *the man* as the object of *understood*, and so would not be maintaining a subject-verb dependency over the following material; see Lewis and Vasishth (2005) for additional discussion. The experiment did not reveal any reading time effects on the critical clause.

2a) The frightened boy understood that the man *who said the townspeople are dangerous* was paranoid about dying.

2b) The frightened boy understood the man *who said the townspeople are dangerous* was paranoid about dying.

2c) The frightened boy understood that the man *who was swimming near the dock* was paranoid about dying.

2c) The frightened boy understood the man *who was swimming near the dock*was paranoid about dying.

Both Hakes et al. (1976) and Van Dyke & Lewis (2003) involve paradigms with relative clauses intervening between subject and verb. However, it is possible that processing of relative clauses involves mechanisms specific to those configurations, and that these null results should not be taken to argue against maintenance cost in subject-verb dependencies more generally. Additionally, Hakes et al. used a phoneme monitoring task, which might not be so informative for syntactic processing, while Van Dyke and Lewis compare clauses with different structures.

Chen et al. (2005) studied whether matrix verb prediction causes processing cost with a different paradigm. They measured self-paced reading times over the clause that was a complement of a verb or noun, and that was identical across conditions. Here, the manipulation concerned the number of verb predictions over the same clause – none, one, or two. They hypothesized that if storage costs are proportional to the number of predicted verbs held in memory, then the zero predicted verbs condition (3a) should be read fastest, and the two predicted verbs condition (3d) the slowest (the predicted verbs are in bold, and the critical region in italics).

3a) Zero predicted verbs

The detective suspected that the thief knew that *the guard protected the jewels* and so he reported immediately to the museum curator.

3b) One late predicted verb

The detective suspected that the knowledge that *the guard protected the jewels* **came** from an insider.

3c) One early predicted verb

The suspicion that the thief knew that *the guard protected the jewels* **worried** the museum curator.

3d) Two predicted verbs

The suspicion that the knowledge that *the guard protected the jewels* **came** from an insider **worried** the museum curator.

The results confirmed their hypothesis - condition (3a) elicited the shortest reading times, while condition (3d) elicited the longest. Thus, Chen et al. (2005) concluded that the parser keeps track of predicted syntactic heads or incomplete subject-verb dependencies, and that there is a cost associated with maintaining these predictions.

In line with this, studies on processing of filler-gap dependencies have frequently found that maintaining a filler increases processing cost on the elements between the filler and the gap site (e.g. King & Just, 1991; Sprouse, Wagers, & Phillips, 2012; Stepanov & Stateva, 2015), which is also confirmed by electrophysiological data (Fiebach et al., 2002; Kluender & Kutas, 1993). Wagers and Phillips (2014) claim that the filler (or at least some information about it) is maintained in the focus of attention during online establishment of this dependency. This result was later confirmed by Ness and Meltzer-Asscher (2017, 2019) and Chow and Zhou (2019).

Interestingly, sentence comprehension models generally assume no maintenance cost related to the processing of matrix subject-verb dependencies, even when prediction or maintenance is an important part of the model's architecture. For example, in Gibson's Syntactic Prediction Locality Theory (SPLT, Gibson, 1998), also known as the Dependency Locality Theory (DLT, Gibson, 2000), the processing cost related to establishing syntactic dependencies is based on the storage cost upon predicting the elements needed to conclude those dependencies, as well as the integration cost once those elements are encountered. However, despite the fact that the dependency between the matrix subject and the corresponding verb in the SPLT model does imply verb prediction, maintaining this prediction is explicitly posited to incur no processing cost, in contrast to maintaining other predictions. This is because a main verb is expected in any clause.

Current study

So far, there is rather mixed evidence as to whether maintaining a subject-verb dependency is costly. Some studies have looked at relative clauses intervening between the subject and the verb in a phoneme monitoring task (Hakes et al., 1976) or in self-paced reading task (Van Dyke & Lewis, 2003), and found no evidence for maintenance cost, while Chen et al. (2005) found evidence of maintenance cost using self-paced reading and complement clauses.

In the current study, we investigate maintenance of the subject-verb dependency using a design with adverbial embedded clauses. In this way, we avoid any additional argument prediction or dependency related to the embedded clause that might be generated in the matrix clause. In addition, we use the eye-tracking technique, which provides more naturalistic data compared to previous paradigms. Our hypothesis is that subject-verb maintenance does increase processing cost on an adverbial clause interpolated between the subject and the verb, which should be manifested as an increase in reading times.

We test our hypothesis in two morphosyntactically different languages: Spanish (Experiment 1) and English (Experiment 2). There are some relevant differences between Spanish and English when it comes to subject-verb dependencies: Verbs are always marked for tense, person and number in Spanish, while in English, they are only marked for tense, and the only person and number marking occurs in present tense, singular number. Thus, the verb in Spanish has to agree with the subject in more features, implying that more subject-related information needs to be available at the verb. This property of Spanish could potentially affect maintenance mechanisms. Indeed, Lago, Shalom, Sigman, Lau, and Phillips (2015) investigated working memory-related effects in English and Spanish subject-verb agreement, and found that the two languages differed regarding the effects on the auxiliary verb. Other studies have shown differences between Spanish and English number agreement processing, such as the role of conceptual factors and the preferred dependency pattern (syntactic or semantic) for certain constructions (Vigliocco, Butterworth, & Garrett, 1996). Given the nature of the subject-verb dependency, as well as the cross-linguistic differences, additional factors were manipulated in the current study to investigate their contribution to the potential maintenance effects: matching between the number marking within the matrix and the embedded clause (Experiments 1 and 2), the number of the matrix subject (Experiment 2), and the word order within the embedded clause (Experiment 1).

Experiment 1

Method

Participants

Fifty-six participants (26 male) took part in the experiment. They were all native Spanish native speakers aged 18 to 41, with the average age of 28.64 (SD=6.96). The experiment was approved by the Basque Center on Cognition, Brain and Language Ethics Review Board and complied with the guidelines of the Helsinki Declaration (World Medical Association, 2013).

Stimuli

The stimuli consisted of 112 Spanish sentences (see Table 1). We refer to the critical conditions as *intervening* and *non-intervening*. The *intervening* sentences were created by embedding an adverbial (temporal or conditional) clause after the subject noun phrase (NP) of a main clause. The main subject NP (matrix NP) consisted of a determiner and a noun only (e.g., *esa chica*, “that girl”). The adverbial clause consisted of a conjunction (e.g., *cuando*, “when”, *cada vez que*, “every time”, or *si*, “if”) followed by a subject NP (e.g., *un chico*, “a boy” / *chicos*, “boys), a verb phrase (VP; e.g., *viene*, “comes”, *vienen*, “come”) and an adverbial phrase. In the *non-intervening* sentences, the same adverbial clause did not interrupt a matrix subject-verb dependency. This was achieved by adding the word *para* (“for”) prior to the matrix subject, thus making it a non-argument NP. Consequently, the rest of the main clause (after the embedded clause) was different.

We also manipulated two additional factors, namely word order within the embedded clause and matching in number between the subjects of the matrix and embedded clauses. These manipulations were added in order to explore different effects that might interact with the factor of interest, e.g. whether the differences in processing difficulty for different word orders (Erdocia, Laka, Mestres-Missé, & Rodriguez-Fornells, 2009; Gattei et al., 2018; Weyerts, Penke, Münte, Heinze, & Clahsen, 2002) affect the temporal dynamics of subject-verb dependency establishment, and, consequently, the maintenance mechanisms involved in it. However, they were not the main focus of the current study and will thus be only briefly discussed. Conditions with VS word order were created by permuting the order of the embedded noun and verb, as illustrated in Table 1.

In the match condition, the adverbial clause subject and verb had the same number as the matrix subject, while in the mismatch condition, the number of the two mismatched. In this experiment, the number of the adverbial clause subject was not manipulated, and was constant across the conditions (see Table 1, underlined), but was balanced across the items (56 singular and 56 plural). Thus, the match/mismatch manipulation was accomplished by varying the number of the matrix subject (and verb) across the conditions, within each item.

The stimuli were Latin squared into 8 lists, and 80 filler items with different structures that do not include embedded clauses were added to each list. The filler items ranged from very simple sentences to more complex sentences with coordinated clauses, but the structures were not intentionally repeated or manipulated. The sentences were created by a native Spanish speaker.

Additionally, a naturalness judgment task was performed to make sure that possible differences in the acceptability of the different word orders (e.g. Erdocia et al., 2009) do not affect our intervention factor. We collected data from 34 native Spanish speakers who did not take part in Experiment 1, two of whom were removed due to their language profile. Their task was to read the sentences and rate whether they sounded natural, on a scale from 0 (=completely unnatural, unusual) to 5 (=completely natural, usual). 128 sentences were Latin Squared into 8 lists, intermixed with 80 filler items used in Experiment 1, and randomized. 16 items were excluded from the analysis (as well as Experiment 1) due to low naturalness. They were presented using PsychoPy software, version 1.83.04 (Peirce et al., 2019). The rating started with 6 practice trials. The overall mean rating was 3.38 (SE=0.01); the intervening mean was 3.49, (SE=0.02), and the non-intervening mean 3.27 (SE=0.02). We fitted linear mixed effects models with the rating score as the dependent variable, and with the intervention, number agreement, and word order as fixed effects, as well as by-item and by-subject random intercepts, and the by-subject random word-order slope. The model selection and the analysis procedure followed the one described in Experiment 1, and the effect coding was applied to the factors. The analysis showed a main effect of intervention, such that intervening condition was rated as more natural [Estimate: 0.23, SE: 0.02, statistic= 8.59], and a main effect of Word Order, such that SV word order was rated as more natural [Estimate: -0.17, SE: 0.04, statistic= -4.29]. No other effects were significant (all ts <1.5).

Procedure

We used an SR EyeLink 1000 eyetracker with a chinrest and the sampling rate of 1000 Hz to record the eye movements (SR Research, Ontario, Canada). Reading was binocular, but only the right eye was tracked. The experiment was presented using Experiment Builder software, version 2.1.45, while the initial stages of data pre-processing were carried out using Data Viewer software, version 2.3.1. The sentences were presented in 19-point Times New Roman font on a black background in one or two lines, depending on the sentence length. There was a 2.5-line space between the two lines. We performed 9-point calibration before the experiment, between the two blocks, and additionally in the case fixation stability was lost. Before each trial, participants had to fixate a red point, positioned on the left side of the screen, where the first word of each item appeared.

Participants were tested individually. They were asked to read the sentences for comprehension and answer the YES/NO comprehension questions that appeared after 50% of the sentences (e.g. for the example in Table 1, the question would be *Is it a boy that was coming to the house in the sentence?*), by pressing keys on the keyboard. The session lasted approximately 45 minutes, and consisted of a practice block (5 items), followed by the two experimental blocks (100 items each).

Data trimming and analysis

Prior to the analysis, a short fixation (≤ 80 ms) was merged with a preceding/following fixation if it was within one character (0.37 degrees); otherwise, these short fixations were removed. The interest area to be analyzed comprised the entire embedded clause (e.g. *Esa chica,[cada vez que viene un chico,] se pone nerviosa*). We calculated first pass and go-past reading times, as well as the regressions out of the interest area to the previous part of the sentence (i.e. the matrix NP). We did not examine total reading time or second pass time, which would reflect re-reading time after the first pass through the relevant material. All the analyses were done using the package *lme4* (Bates, Machler, Bolker, & Walker, 2015) in *R* Statistical software (R Core Team, 2017). We fitted linear mixed-effect models (for reading times) or generalized linear mixed-effect models (for proportion of regressions), with intervention (non-intervening, intervening), number agreement (match, mismatch) and word order (SV, VS) as fixed effects. We tested a series of models for exclusion of random by-item and by-subject slopes (by-item and by-subject random intercepts were left by default, except when resulting in a non-converging model, in which case by-subject intercepts were kept), using a backward-direction model selection method (Barr, Levy, Scheepers, & Tily, 2013). The slopes were included when falling below the model-selection α level, which was .05 in this study, and if test didn't fall below the model-selection α level, then the simpler model was retained. When encountering non-convergence or singular fit problems, the random effects structure was progressively simplified (by removing random slopes – by-item and by-participant) until reaching convergence. Effect coding was used for all the factors (for the intervention factor: -0.5= non-intervening, 0.5 = intervening; for the number agreement factor: -0.5 = match, 0.5 = mismatch condition; for the word order factor: -0.5 = SV word order, 0.5 = VS word order condition). We report the intercept, the estimate, standard error, and the t/z value in tables. An effect was considered significant when *t/z* >= 2. All the analyses are available in the online supplemental materials (Ristic, 2020).

Results

The mean accuracy on comprehension questions in Experiment 1 was 91.68% (SD=5.3). All participants scored between 68.75% and 96.87%.

For first pass reading time, the only significant effect was a main effect of Word Order, driven by longer first pass reading times for the VS word order. There was also a marginal Number Agreement x Word Order interaction, as the VS word order elicited longer reading times in the matching condition. Also, there was a marginal Intervention x Number Agreement x Word Order interaction, as VS word order elicited longer reading times in matching conditions in the intervening items (see Figure 1 and Table 2).

For go-past reading time, longer reading times for the VS word order generated a significant effect of Word Order, while the increase in the go-past reading times for the intervening condition generated a significant main effect of intervention (see Figure 1 and Table 2).

The analysis for the regressions out of the embedded clause to the previous interest areas showed only a significant main effect of intervention, as more regressions were made in the

Intervening items. There was also a marginal effect of word order, driven by more regressions in VS word order, as well as a marginal effect of number agreement, as more regressions emerged in the mismatch condition. Finally, we found a marginal Word Order x Number Agreement interaction, driven by an increase in regressions in VS word order for the mismatch items (see Figure 1 and Table 2).

Discussion

The findings from Experiment 1 support our main hypothesis: Maintaining a subject-verb dependency in sentence comprehension does induce a processing cost. Both go-past reading times on the adverbial clause and regressions out of this clause to the previous regions were significantly increased in the intervening condition (i.e. when the embedded clause interrupts a subject-verb dependency) compared to the non-intervening condition (i.e. when the same clause doesn't interrupt the subject-verb dependency). Note that in the naturalness judgements, the items where the intervention was present received higher ratings, not lower, corroborating our claim that the eye-tracking effects come from matrix subject-verb dependency maintenance, and not general difficulty in processing these sentences.

There was also a significant main effect of Word Order on both first-pass and go-past reading times, with longer reading times with VS word order. This result is consistent with previous research (Erdocia et al., 2009; Gattei et al., 2018; Weyerts et al., 2002), and was further confirmed by our naturalness judgment rating, where SV word order was rated as more natural.

In Experiment 2, we attempt to replicate in English the central finding of Experiment 1, i.e. increased reading times on an adverbial clause that intervenes in a subject-verb dependency.

Experiment 2

Method

Participants

Fifty six participants (10 male, 46 female) aged 18 to 23 (mean age = 19.69, SD = 1.02) took part in the experiment for a course credit. All were native English speakers. The experiment was approved by the University of Massachusetts Amherst Institutional Review Board.

Stimuli

The stimuli consisted of 80 English sentences (see Table 3), similar to those in Experiment 1. However, one of the two additional morphosyntactic factors was different in Experiment 2. Word order in the embedded clause wasn't manipulated in Experiment 2, given that English doesn't allow for VS word order. Instead, both the number of the matrix subject NP (singular/plural) and the number of the embedded NP (singular/plural) were manipulated. Like in Experiment 1, experimental sentences were created by embedding an adverbial (temporal or conditional) clause after the subject NP of a main clause (matrix NP). In the intervening condition, the matrix NP consisted of a determiner and a noun (*that student/ those students*). The embedded clause consisted of a conjunction (*when, whenever, every time, as soon as, if*), followed by an embedded NP (*a professor/ professors*), an embedded VP (*finishes/ finish*) and an adverbial. As in Experiment 1, we additionally manipulated matching in number between the matrix and the embedded clause. In the match condition, the embedded NP/VP had the same number as the matrix NP, while in the mismatch condition, the number of the embedded NP/VP was different from the number of the matrix NP; the adverbial subject NP was manipulated to be either singular or plural. Again, these two factors were added in order to further explore morphosyntactic effects that might interact with the intervention factor, but they do not directly relate to the current study's main hypothesis and will only be briefly discussed.

The non-intervening items were different than those in Experiment 1, as all the obligatory matrix clause elements were placed before the embedded clause. By placing all the obligatory matrix clause arguments before the embedded clause in the non-intervening items, we made sure to eliminate the possibility (present in Experiment 1) that some predictions regarding the matrix clause are maintained even in the non-intervening items. To create the English non-intervening items, an initial subject and verb were added so that the NP that served as the matrix subject in the intervening items became a direct or indirect object. The rest of the sentence (after the adverbial clause) was the same as in the intervening items, except for an additional pronoun (*she*; see Table 3). Also, the conjunction *and* was added at the beginning of the embedded clause in the non-intervening items, but it was not included in the analyzed region. The sentences were created by 3 native English speakers. We used a Latin square design to divide the stimuli into 8 lists, to which 80 filler items with different structures that did not include intervening clauses were added. Some filler items contained adverbial phrases or adverbial clauses, but they were different from those in the experimental items, and they were never located between the matrix subject and verb. The full set of stimuli sentences can be found in the online supplemental materials (Ristic, 2020).

Procedure

We used an SR EyeLink 1000 eyetracker with the chinrest and the sampling rate of 1000 Hz to record eye movements (SR Research, Ontario, Canada). Reading was binocular, but only the right eye was tracked. The sentences were presented in 11-point Monaco font on a white background, in one line. We performed 3-point calibration before the experiment, and additionally if the fixation stability was lost. The participants triggered each sentence by fixating a box at the left edge of the monitor. The experiment was carried out using the EyeTrack software, and initial stages of data analysis were performed with Robodoc and EyeDry (http://blogs.umass.edu/eyelab/software/).

Participants were tested individually. They were asked to read for comprehension and answer YES/NO comprehension questions that appeared after 50% of the sentences, by pressing buttons on a console. The questions were similar to those in Experiment 1. The session consisted of a practice block (8 items), followed by the experimental block, which lasted approximately 40 minutes.

Data trimming and analysis

We fitted linear mixed-effect models (for reading times) or generalized linear mixed-effect models (for proportion of regressions), with intervention, number agreement, and embedded NP number as fixed effects. Effect coding was used for all the factors (for the intervention factor: -0.5= non-intervening, 0.5 = intervening; for the number agreement factor: -0.5 = match, 0.5 = mismatch condition; for the embedded NP number factor: -0.5 = plural, 0.5 = singular condition). The rest of the data trimming and analysis followed the procedures described in Experiment 1. All the analyses are available in the online supplemental materials (Ristic, 2020).

<u>Results</u>

The mean accuracy on comprehension questions in Experiment 2 was 90.80% (SD=6.5). All participants scored between 70% and 100%.

In first pass reading time, we found a marginal effect of intervention, driven by an increase in the first pass reading times for the intervening condition. Also, the effect of embedded NP number was significant, as the singular embedded clauses elicited longer first pass reading times (see Figure 2 and Table 4).

In the go-past reading time analysis, we observed the same effects and in the same directions as those in the first pass reading times. However, the effect of intervention was significant, while the effect of embedded NP number was marginal (see Figure 2 and Table 4).

No significant effects emerged for regressions out (all *z*s< 1.00, see Figure 2 and Table 4).

Discussion

The analysis of the adverbial clause region in Experiment 2 showed comparable effects to those found in Experiment 1: we found a significant effect of intervention, such that reading times increased when this clause was located between the matrix subject and the matrix verb. This supports our finding that establishing a subject-verb dependency in sentence comprehension entails maintenance cost. The effect of intervention in Experiment 2 was significant in go-past reading times, and marginally significant in first pass reading times. However, although numerically present, it was not significant in the regressions out of the adverbial clause to the previous regions.

Furthermore, we also found that the adverbial clauses containing singular subjects and verbs were somewhat harder to process, eliciting longer first pass and go-past reading times. However, neither this effect nor the effect of number agreement interacted with the effect of intervention.

General Discussion

In the current study, we set out to explore whether establishing a subject-verb dependency results in increased processing cost on elements between the subject and the verb. Both the English and the Spanish experiment showed that there is an increased processing cost on elements located between the matrix subject and the matrix verb. Like Chen et al. (2005), we measured reading times on embedded (non-relative) clauses that are identical across conditions. However, we used adverbial embedded clauses, making sure that no syntactic expectations or relations are established between the matrix and embedded clause. In line with the findings of Chen et al. (and contra Hakes et al., 1976 and Van Dyke & Lewis, 2003), we show that there's an increase in reading times over the interpolated elements, confirming that subject-verb dependency establishment incurs maintenance cost. Therefore, we provide the first eye-tracking evidence for maintenance cost related to subject-verb dependency establishment in sentence comprehension. Moreover, our results show that this effect is present cross-linguistically, as it occurred both in English and Spanish.

In both experiments, the effect of intervention seems to be most prominent in go-past reading times and regressions out (although the effect on regressions out was not statistically significant in Experiment 2). The increase in re-reading and regressions from an intervening clause might reflect the difficulty of processing this material while maintaining information. That is, processing is harder when additional resources are spent to maintain information related to the establishment of a subject-verb dependency, and therefore elements in the embedded clause require more time and more regressions. In the non-intervening condition, no resources are spent to maintain information as no obligatory dependency establishment is being interrupted.

It is also important to highlight that the effect in go-past reading time might be stronger than it appears. In both experiments, the materials are designed such that longer go-past reading times might be expected in the non-intervening condition, compared to the intervening condition, in the absence of any effect of dependency maintenance. This is because the material preceding the adverbial clause is longer in the non-intervening items, and therefore there is more text to re-read. Finding effects in the opposite direction strengthens our claim that the increase in the intervening condition reflects maintenance cost. Furthermore, it should be highlighted that the non-intervening items were different across the experiments, such that some argument prediction could have been potentially maintained over the adverbial clause in the non-intervening items of Experiment 1. Still, the intervention effect was present in both experiments, suggesting that it can be specifically ascribed to the matrix subject-verb dependency, and not to general argument prediction maintenance.

Together with the findings by Chen et al. (2005), our results question the assumption of the SPLT/DLT model that matrix subject-verb prediction is maintained, but causes no processing cost (Gibson, 1998, 2000). We therefore believe that the SPLT/DLT model should accommodate these findings. Specifically, additional storage cost could be added to each element that interpolates between the subject and the verb, reflecting the cost of maintenance of representation(s) during the subject-verb dependency establishment.

It is also interesting to discuss our results from the perspective of another well-known sentence comprehension model, the retrieval-based model of Lewis and Vasishth (2005). This model attributes the processing cost found in center-embedded clauses to difficulty occurring at retrieval, when a dependency is completed. That is, no storage-load effects are assumed as no processing effort is dedicated to keep prior constituents (and thus matrix subject and verb prediction) active. Lewis and Vasishth (2005) explicitly reference the lack of apparent maintenance cost in the Van Dyke and Lewis (2003) study, discussed above, as support for the notion that when retrieval cost is equated, there is no cost associated with dependency maintenance. However, in the current study, there are no apparent differences in retrieval requirements in the intervening and non-intervening conditions. The current study suggests that matrix subject-verb dependencies might need to be kept in a privileged state, and that contrary to the assumptions of the Lewis and Vasishth (2005) model, this does induce some cost.

Our findings do have certain limitations. As noted above, the two studies presented here used slightly different control (non-intervening) items. Although we believe that this does not diminish our conclusions, a study with more similar non-intervening items might have provided a closer cross-linguistic comparison. Furthermore, the additional manipulated factors also differed across the studies, due to the nature of the languages. Certainly, the word order effects in the Spanish experiment could have been better understood with a more careful manipulation of the verb types used across the items, as different verbs (e.g. unaccusative/unergative) might show preference for different word orders[1]. Nonetheless, we believe that the results arising from these additional factors (and their potential cross-linguistic differences) hint at possible questions to be addressed in future studies.

For example, an interesting question concerns the nature of the maintained representations, and future research should focus on determining both their number and composition. Research on filler-gap dependencies suggests that the information about the category (Wagers & Phillips, 2014) as well as semantics (Chow & Zhou, 2019; Ness & Meltzer-Asscher, 2019) of the filler element can form part of the maintained representation. Previous research on subject-verb dependency processing (Chen et al., 2005) suggests that the matrix verb prediction is maintained, but it is still unclear whether the verb category is the only maintained information. The current study suggests, but does not clearly establish, that number information might be relevant, and that the relevance of number information might differ cross-linguistically. Certainly, it would be interesting to investigate this issue further, as it has already been suggested that morphosyntactic information such as number is processed in a top-down, predictive manner in subject-verb dependencies (Lago et al., 2015; Tanner, Nicol, & Brehm, 2014).

In conclusion, the current study provides the first eye-tracking data showing a maintenance cost in long-distance subject-verb dependency establishment. Arguably, this maintenance is purposeful and intentional, as its final goal is a successful and efficient completion of the dependency. Finally, future research should look into the nature of the maintained representation,

[1] We thank an anonymous reviewer for this observation.

and sentence comprehension models should account for the maintenance cost generated by matrix subject-verb dependency processing.

Acknowledgements

The authors thank Ainhoa Eguiguren, Amber Hughes and Colin McDermott for their help with stimuli preparation. This research was partially funded by the following grants: PRE_2018_2_0074 and EP_2018_1_0042 (B.R.), as well as PI_2016_1_0014 (N.M.) by Eusko Jaurlaritza, PSI2015-65694-P and RTI2018-096311-B-I00 to N.M., RYC-2017-22015 and FFI2016-76432-P_LAMPT to S.M., funded by the Ministerio de Economía y Competitividad, Agencia Estatal de Investigación & Fondo Europeo de Desarrollo Regional, as well as through Severo Ochoa excellence program to BCBL (SEV-2015-0490) by Agencia Estatal de Investigación.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing : Keep it maximal. *Journal of Memory and Language*, *44*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language*, *52*(1), 144–169. https://doi.org/10.1016/j.jml.2004.10.001

Chow, W. Y., & Zhou, Y. (2019). Eye-tracking evidence for active gap-filling regardless of dependency length. *Quarterly Journal of Experimental Psychology (2006)*, *72*(6), 1297–1307. https://doi.org/10.1177/1747021818804988

Erdocia, K., Laka, I., Mestres-Missé, A., & Rodriguez-Fornells, A. (2009). Syntactic complexity and ambiguity resolution in a free word order language: Behavioral and electrophysiological evidences from Basque. *Brain and Language*, *109*(1), 1–17. https://doi.org/10.1016/j.bandl.2008.12.003

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German WH-questions. *Journal of Memory and Language*, *47*, 250–272. https://doi.org/10.1016/S0749-596X(02)00004-9

Gattei, C. A., Sevilla, Y., Tabullo, Á. J., Wainselboim, A. J., París, L. A., & Shalom, D. E. (2018). Prominence in Spanish sentence comprehension: an eye-tracking study. *Language, Cognition and Neuroscience*, *33*(5), 587–607. https://doi.org/10.1080/23273798.2017.1397278

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

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, Language, Brain*, 95–126. https://doi.org/10.7551/mitpress/3654.003.0008

Hakes, D. T., Evans, J. S., & Brannon, L. L. (1976). Understanding sentences with relative clauses. *Memory & Cognition*, *4*(3), 283–290. https://doi.org/10.3758/BF03213177

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602. https://doi.org/10.1016/0749-596X(91)90027-H

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, *5*(2), 196–214. https://doi.org/10.1162/jocn.1993.5.2.196

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, *82*, 133–149. https://doi.org/10.1016/j.jml.2015.02.002

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

Ness, T., & Meltzer-Asscher, A. (2017). Working Memory in the Processing of Long-Distance Dependencies: Interference and Filler Maintenance. *Journal of Psycholinguistic Research*, *46*(6), 1353–1365. https://doi.org/10.1007/s10936-017-9499-6

Ness, T., & Meltzer-Asscher, A. (2019). When is the verb a potential gap site? The influence of filler maintenance on the active search for a gap. *Language, Cognition and Neuroscience*, *34*(7), 936–948. https://doi.org/10.1080/23273798.2019.1591471

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. Behavior Research Methods. 10.3758/s13428-018-01193-y

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*(3), 407–428. https://doi.org/10.1016/j.cogbrainres.2004.09.012

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ristic, B. (2020). *Maintenance cost in the processing of subject-verb dependencies* [Data sets and scripts]. OSF. https://doi.org/10.17605/OSF.IO/VS6C5

Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working-memory capacity and syntactic island effects. *Language*, *88*(1), 82–123. https://doi.org/10.1353/lan.2012.0004

SR Research. (n.d.). SR Research Experiment Builder. Ontario,Canada.

Stepanov, A., & Stateva, P. (2015). Cross-linguistic evidence for memory storage costs in filler-gap dependencies with wh-adjuncts. *Frontiers in Psychology*, *6*(September), 1–19. https://doi.org/10.3389/fpsyg.2015.01301

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, *76*, 195–215. https://doi.org/10.1016/j.jml.2014.07.003

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316. https://doi.org/10.1016/S0749-596X(03)00081-0

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition*, *61*(3), 261–298. https://doi.org/10.1016/S0010-0277(96)00713-5

Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology*, *67*(7), 1274–1304. https://doi.org/10.1080/17470218.2013.858363

Weyerts, H., Penke, M., Münte, T. F., Heinze, H. J., & Clahsen, H. (2002). Word order in sentence processing: An experimental study of verb placement in German. *Journal of Psycholinguistic Research*, *31*(3), 211–268. https://doi.org/10.1023/A:1015588012457

World Medical Association. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. Jama, 310(20), 2191-2194.

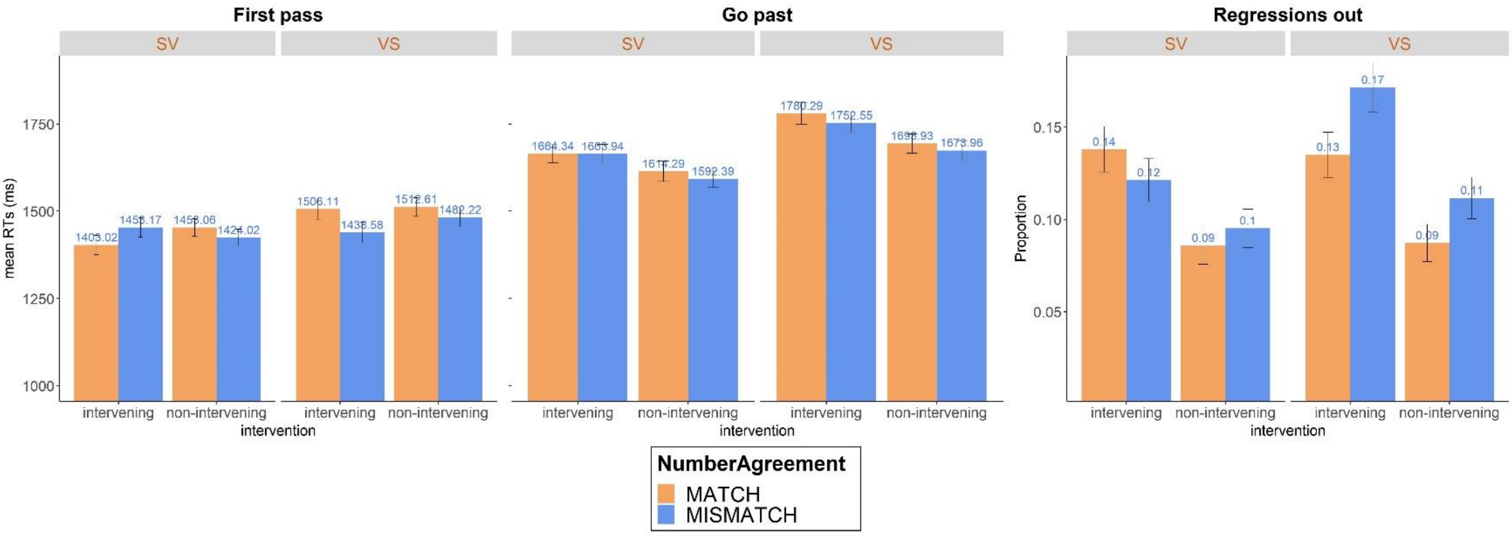

Figure 1. *First Pass Reading Times, Go-Past Reading Times and Regressions Out of the Embedded Clause in Experiment 1*

Note. The numbers on the top of the bars represent the mean values for that level of the factor, and error bars indicate the standard error of the mean. SV = subject–verb word order; VS = verb–subject word order; RT = reaction time. See the online article for the color version of this figure.

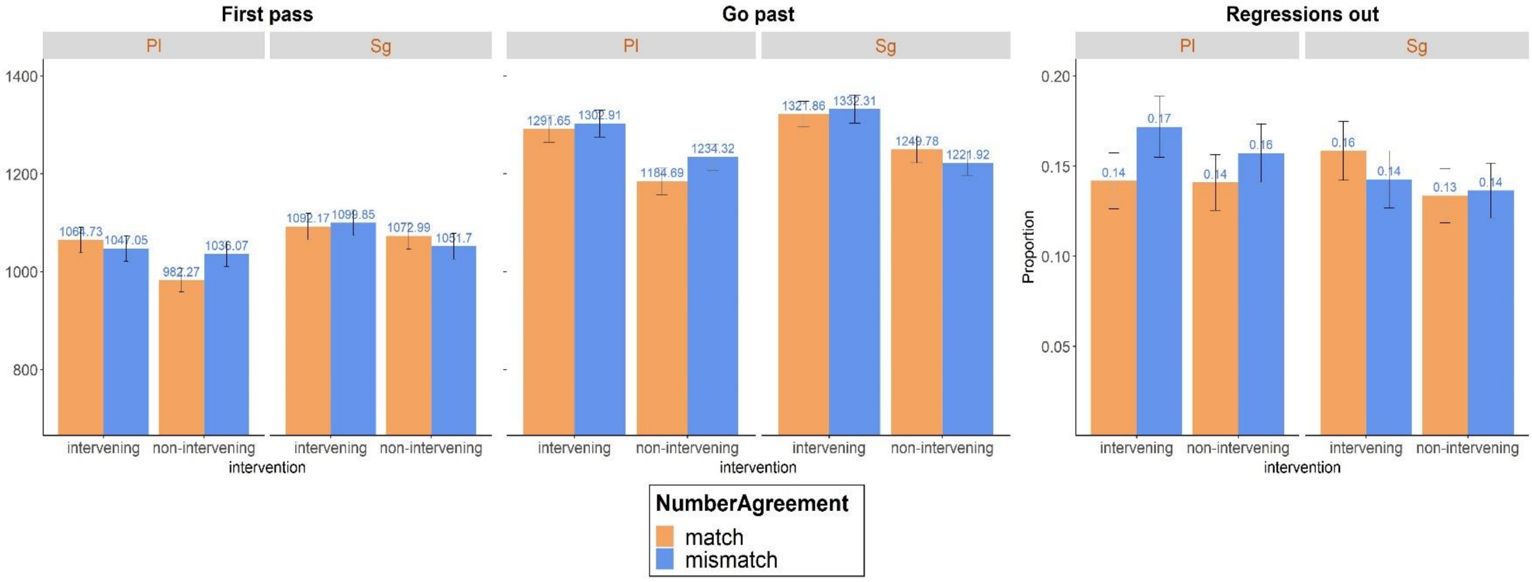

Figure 2. *First Pass Reading Times, Go-Past Reading Times, and Regressions Out of the Embedded Clause in Experiment 2*

Note. The numbers on the top of the bars represent the mean values for that level of the factor, and error bars indicate the standard error of the mean. NP = number and person; Pl = plural embedded NP; Sg = singular embedded NP. See the online article for the color version of this figure.

Word order	Intervention	Number agreement	Example sentence[a]
Subject-verb	intervening	match	Esa chica, **cada vez que un chico viene a casa,** se pone nerviosa. “That girl, every time a boy comes to the house, gets nervous.”
		mismatch	Esas chicas, **cada vez que un chico viene a casa,** se ponen nerviosas. “Those girls, every time a boy comes to the house, get nervous.”
	non-intervening	match	Para esa chica, **cada vez que un chico viene a casa**, todo lo demás es aburrido. “To that girl, every time a boy comes to the house, everything else is boring.”
		mismatch	Para esas chicas, **cada vez que un chico viene a casa,** todo lo demás es aburrido. “To those girls, every time a boy comes to the house, everything else is boring.”
Verb-subject	intervening	match	Esa chica, **cada vez que viene un chico a casa,** se pone nerviosa. “That girl, every time a boy comes to the house, gets nervous.”
		mismatch	Esas chicas, **cada vez que viene un chico a casa,** se ponen nerviosas. “Those girls, every time a boy comes to the house, get nervous.”
	non-intervening	match	Para esa chica, **cada vez que viene un chico a casa**, todo lo demás es aburrido**.** “To that girl, every time a boy comes to the house, everything else is boring.”
		mismatch	Para esas chicas, **cada vez que viene un chico a casa**, todo lo demás es aburrido. “To those girls, every time a boy comes to the house, everything else is boring.”

Table 1. *Example of an Experimental Item with Singular Embedded NP in Experiment 1*

[a]The embedded clause is in bold, with the embedded noun and verb underlined.

***First pass* reading times**

model: firstpass ~ Intervention X NumberAgreement X WO + (1 | participant) + (1 | item)					
	estimate	std.error	statistic	conf.low	conf.high
(Intercept)	1458.58	60.23	24.22	1340.53	1576.63
Intervention1	-16.74	15.59	-1.07	-47.29	13.81
NumberAgreement1	-18.07	15.59	-1.16	-48.62	12.49
WO1	51.81	15.59	**3.32**	21.25	82.36
Intervention1 X NumberAgreement1	23.53	31.18	0.75	-37.57	84.64
Intervention1 X WO1	-12.52	31.18	-0.40	-73.62	48.59
NumberAgreement1 X WO1	-58.24	31.18	-1.87	-119.35	2.87
Intervention1 X NumberAgreement1 X WO1	-114.62	62.36	-1.84	-236.83	7.60

***Go-past* reading times**

model: gopast ~ Intervention X NumberAgreement X WO + (1| participant) + (1| item)					
	estimate	std.error	statistic	conf.low	conf.high
(Intercept)	1679.31	71.75	23.41	1538.69	1819.93
Intervention1	72.42	13.67	**5.30**	45.64	99.21
NumberAgreement1	-16.02	13.67	-1.17	-42.80	10.77
WO1	91.84	17.47	**5.26**	57.60	126.08
Intervention1 X NumberAgreement1	9.18	27.33	0.34	-44.39	62.75
Intervention1 X WO1	24.63	27.33	0.90	-28.94	78.20
NumberAgreement1 X WO1	-10.68	27.33	-0.39	-64.25	42.89
Intervention1 X NumberAgreement1 X WO1	-28.37	54.66	-0.52	-135.51	78.77

Regressions out

model: regout ~ Intervention X NumberAgreement X WO + (1|participant) + (1|item)						
	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-2.35	0.13	-17.99	0.00	-2.61	-2.10
Intervention1	0.49	0.08	**5.95**	0.00	0.33	0.65
NumberAgreement1	0.14	0.08	1.76	0.08	-0.02	0.31
WO1	0.16	0.08	1.95	0.05	0.00	0.32
Intervention1 X NumberAgreement1	-0.13	0.16	-0.80	0.42	-0.45	0.19
Intervention1 X WO1	0.11	0.16	0.66	0.51	-0.21	0.43
NumberAgreement1 X WO1	0.32	0.16	1.95	0.05	0.00	0.64
Intervention1 X NumberAgreement1 X WO1	0.31	0.33	0.94	0.35	-0.34	0.95

Table 2. *Analyses for the Embedded Clause in Experiment 1*

Note. Coding is as follows: For the intervention factor, -0.5 = nonintervening, 0.5 = intervening; for the number agreement factor, -0.5 = match, 0.5 = mismatch; for the word order factor, -0.5 = subject–verb word order, 0.5 = verb–subject word order. Significant effects are in bold. WO = word order; regout = regressions out.

Intervention	Embedded NP number	Number agreement	Example sentence[a]
intervening	Singular	match	That student, **as soon as a professor finishes the class**, leaves the classroom.
		mismatch	Those students, **as soon as a professor finishes the class**, leave the classroom.
	Plural	match	Those students, **as soon as professors finish the class**, leave the classroom.
		mismatch	That student, **as soon as professors finish the class**, leaves the classroom.
non-intervening	Singular	match	I watched that student, **and as soon as a professor finishes the class**, she leaves the classroom.
		mismatch	I watched those students, **and as soon as a professor finishes the class**, they leave the classroom.
	Plural	match	I watched those students, **and as soon as professors finish the class**, they leave the classroom.
		mismatch	I watched that student, **and as soon as professors finish the class**, she leaves the classroom.

Table 3. *Example of an Experimental Item in Experiment 2*

Note. NP = noun phase.

[a] The embedded clause is in italics, and the embedded noun and verb are underlined.

***First pass* reading times**

model: firstpass~ Intervention X EmbeddedNPNumber X NumberAgreement + (1+intervention|participant) + (1+intervention|item)

	estimate	std.error	statistic	conf.low	conf.high
(Intercept)	1063.56	42.90	24.79	979.47	1147.64
Intervention1	42.83	24.51	1.75	-5.21	90.87
NumberAgreement1	6.20	15.17	0.41	-23.54	35.94
EmbeddedNPNumber1	44.09	15.16	**2.91**	14.37	73.81
Intervention1 X NumberAgreement1	-22.86	30.35	-0.75	-82.35	36.63
Intervention1 X EmbeddedNPNumber1	-12.32	30.34	-0.41	-71.79	47.15
NumberAgreement1 X EmbeddedNPNumber1	-21.77	30.35	-0.72	-81.25	37.71
Intervention1 X NumberAgreement1: EmbeddedNPNumber1	102.19	60.74	1.68	-16.86	221.25

***Go-past* reading times**

model: gopast ~ Intervention X EmbeddedNPNumber X NumberAgreement+ (1|participant) + (1|item)

	estimate	std.error	statistic	conf.low	conf.high
(Intercept)	1276.37	54.84	23.27	1168.89	1383.86
Intervention1	92.07	13.82	**6.66**	64.99	119.15
NumberAgreement1	11.64	13.80	0.84	-15.41	38.68
EmbeddedNPNumber1	25.17	13.79	1.82	-1.86	52.20
Intervention1 X NumberAgreement1	0.18	27.61	0.01	-53.93	54.29
Intervention1 X EmbeddedNPNumber1	0.95	27.61	0.03	-53.17	55.06
NumberAgreement1 X EmbeddedNPNumber1	-41.09	27.59	-1.49	-95.16	12.99
Intervention1 X NumberAgreement1 X EmbeddedNPNumber1	75.28	55.26	1.36	-33.03	183.59

Regressions out

model: regout ~ Intervention X EmbeddedNPNumber X NumberAgreement + (1 |participant) + (1|item)

	estimate	std.error	statistic	conf.low	conf.high
(Intercept)	0.15	0.01	12.01	0.12	0.17
Intervention1	0.01	0.01	1.01	-0.01	0.03
NumberAgreement1	0.01	0.01	0.80	-0.01	0.03
EmbeddedNPNumber1	-0.01	0.01	-0.91	-0.03	0.01
Intervention1 X NumberAgreement1	0.00	0.02	-0.04	-0.04	0.04
Intervention1 X EmbeddedNPNumber1	0.01	0.02	0.30	-0.04	0.05
NumberAgreement1 X EmbeddedNPNumber1	-0.03	0.02	-1.36	-0.07	0.01
Intervention1 X NumberAgreement1 X EmbeddedNPNumber1	-0.03	0.04	-0.73	-0.12	0.05

Table 4. *Analyses for the Embedded Clause in Experiment 2*

Note. Coding is as follows: For the intervention factor, -0.5 = nonintervening, 0.5 = intervening; for the number agreement factor, -0.5 = match, 0.5 = mismatch; for the embedded NP number factor, -0.5 = plural, 0.5 = singular. Significant effects are in bold. NP = noun phrase; gopast = go past; regout = regressions out.