

# Processing syntactic and semantic information in the L2: Evidence for differential cue-weighting in the L1 and L2

Nazik Dinçtopal Deniz 

Boğaziçi University, Department of Foreign Language Education, İstanbul, Turkey

## Research Article

**Cite this article:** Deniz ND (2022). Processing syntactic and semantic information in the L2: Evidence for differential cue-weighting in the L1 and L2. *Bilingualism: Language and Cognition* **25**, 713–725. <https://doi.org/10.1017/S1366728921001140>

Received: 28 January 2021

Revised: 5 December 2021

Accepted: 7 December 2021

First published online: 10 March 2022

**Keywords:**

second language processing; cue-retrieval; cue-weighting; syntax; semantics

**Address for correspondence:**Nazik Dinçtopal Deniz,  
Boğaziçi University, Department of Foreign  
Language Education, Bebek, 34342, İstanbul,  
Turkey  
E-mail [nazik.dinctopal@boun.edu.tr](mailto:nazik.dinctopal@boun.edu.tr)**Abstract**

This study investigated how second language speakers use syntactic and semantic cues in processing complex sentences. Turkish speakers of English, native speakers of English and native speakers of Turkish participated in a self-paced reading experiment, a read-aloud task and a pen-and-paper questionnaire in their relevant languages. The participants' working memory capacity was also measured. The results supported the primacy of syntax view for native speakers of Turkish and English (Frazier & Fodor, 1978). Both groups of native speakers primarily used syntactic cues in their on-line decisions; semantic information influenced later decisions only. The second language speakers, however, used semantic cues in both their initial and later decisions, with evidence for accessing complex syntactic representations. The results are taken to support the view that native and non-native speakers weight linguistic cues differentially (Cunnings, 2017). The potential reasons for differential cue-retrieval in the first and second language are discussed.

**Introduction**

There has been much research reporting discrepancies between first language (L1) and second language (L2) processing. Whether such discrepancies provide evidence for qualitative processing differences between L1 and L2 has been a focus of debate. Several studies contributing to this debate examined whether or not L2 speakers use syntactic information similarly to L1 speakers. But some other studies revealed that L2 speakers rely more heavily on semantic information than L1 speakers in their parsing decisions (Dinçtopal-Deniz, 2010; Grüter, Lau & Ling, 2020).

The increased reliance on semantic information in the L2 has, so far, only been addressed under the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006a, 2006b, 2006c). Clahsen and Felser argue that “L2 processing may prioritize semantic, pragmatic, or other types of nongrammatical information, with L2 speakers potentially being more sensitive to these types of information compared to L1 speakers” and that “this part of [their] hypothesis has rarely been investigated” (2018, p. 695). But in the SSH L2 speakers are also not expected to compute complex hierarchical representations. In SSH “[l]earners appear to be capable, in principle, of processing grammatical phenomena involving locally related constituents in a native-like fashion but not those involving structurally complex phenomena such as nonlocal dependencies, indicating that L2 learners do not have problems with all aspects of grammar but with the real-time computation of complex hierarchical representations.” (Clahsen & Felser, 2006c, p. 568).

An alternative view by Cunnings (2017) maintains that L2 speakers do not differ from L1 speakers in their access to syntactic representations but they diverge in how they retrieve cues from memory. According to Cunnings, L2 speakers are sensitive to syntactic information and “do construct fully-specified syntactic parses, but ... they are more susceptible to interference during memory retrieval than L1 speakers.” (p. 662) and “... [they] rely more heavily on discourse-based cues to memory retrieval than L1 speakers.” (p. 663). Although Cunnings compares weighting of discourse-related cues to that of syntactic cues, there may also be discrepancy between L1 and L2 weighting of semantic versus syntactic cues given the previous findings for increased reliance on semantic information in the L2 (e.g., Grüter et al., 2020).

The present study tests whether a shallow parsing view or differential cue-weighting in the L2 can better explain L2 speakers' sentence processing. Experimental sentences include relative clause (RC) attachment ambiguities such as *The servant of the actress who was on the balcony was shot*. The structure is ambiguous as to which noun phrase (NP) *the servant* or *the actress*, the RC, *who was on the balcony*, would attach to. Processing RC attachment ambiguities requires computation of complex syntactic hierarchy as the RC is headed by a complex NP. The present study disambiguates the attachment site via semantic information, more specifically the animacy of the NPs in the complex genitive subject (details in the Present Study

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Section). An SSH-based approach would predict no reliable RC attachment preference in the L2 speakers' on-line parsing decisions for such constructions as complex syntactic detail would not be available to them despite its prediction of increased reliance on semantic information. Cunnings' approach, however, would predict that the full parsing route with syntactic detail would be available to the L2 learners but their on-line parsing decisions could be different from those by L1 learners due to their differential cue retrieval from memory. Thus, if the L2 speakers of the present study show a clear attachment preference and if such preferences are modulated by semantic information, that would suggest heavier weighing of non-syntactic cues in the L2. A third alternative is native-like L2 parsing. That observation would refute both views.

The present study also tests capacity-based approaches to L2 parsing. Dinçtopal-Deniz (2010) reported, in a smaller scale study, that Turkish learners of English preferred to attach the RC high in their English off-line decisions. Since Turkish, similar to English, shows low-attachment preference (Dinçtopal-Deniz, 2010; Kirkic, 2004) an L1-based parsing strategy could not explain the finding. But it is possible that L2 speakers' working memory (WM) capacity or prosodic phrasings may provide an explanation.

Swets, Desmet, Hambrick, and Ferreira (2007) found for English and Dutch L1 speakers that low-span readers had a tendency to attach RCs high whereas high-span readers preferred low attachment. This was attributed to their chunking strategies. Low-span readers, preferring smaller chunks, would form shorter phrases and insert an implicit prosodic boundary before the RC, leading to high attachment decisions. High-span readers would process the RC and the complex NP within the same prosodic phase, resulting in low-attachment preferences.

It has also been suggested that processing in the L2 increases the demands on WM (Hopp, 2014; Temple, 2000) and WM capacity may be smaller in the L2 (Harrington & Sawyer, 1992). If so, the L2 speakers in Dinçtopal-Deniz (2010) may have behaved similar to low WM capacity native speakers (see Hopp, 2014 for a similar argument) and attached the RC high due to forming smaller chunks (Swets et al., 2007). Thus, the present study also investigates the participants' WM capacity as a potential predictor of RC attachment preference and examines their prosodic phrasing to test if it influences their parsing decisions and if these decisions are modulated by WM capacity.

### *RC attachment in L2 processing*

In one of the first bilingual studies on RC attachment ambiguity resolution Fernández (2002) examined RC attachment preferences of Spanish–English bilinguals. Neither the English-dominant nor Spanish-dominant bilinguals in the study showed any reliable attachment preference in the self-paced reading task in either language. The off-line data for bilinguals showed a consistent effect of language dominance: the Spanish-dominant group was more likely to attach high in both languages than the English-dominant group. Fernández took this as evidence for language independent (universal) processing strategies in bilinguals.

Dussias (2003) investigated RC attachment preferences of L1 Spanish–L2 English and L1 English–L2 Spanish speakers. In her study, both groups showed a low-attachment preference in both languages. Dussias attributed this to the cognitive demand placed on the bilingual processor and to language exposure. Since hosting two languages would be cognitively costlier, the bilingual

processor presumably opted for a memory-friendly option by attaching local. It was also possible that the bilinguals' RC attachment resembled that of English (i.e., low attachment) since they lived in an English-speaking environment.

Papadopoulou and Clahsen (2003) and Felser, Roberts, Marinis, and Gross (2003b) examined the role of thematic and syntactic information in RC attachment ambiguity resolution, with L2 Greek and L2 English speakers, respectively. The constructions involved genitive (NP-*of*-NP) or PP antecedents (NP-*with*-NP). When there was a theta-role-assigning lexical preposition such as *with*, the non-native speakers, from all L1 backgrounds, patterned similar to natives (i.e., a low-attachment preference), but when the whole complex DP was the local thematic processing domain (i.e., a genitive antecedent), they, unlike native speakers, did not show any attachment bias. The results were taken to suggest that, although L1 parsing is informed by both syntactic and lexical information, L2 parsing is mainly guided by lexical information.

Omaki (2005) investigated the role of WM capacity in L1 Japanese–L2 English speakers' RC attachment preferences. Off-line attachment preferences of English speakers were influenced by WM capacity (low attachment for high-span, no preference for low-span); the self-paced reading data showed no effect of reading-span. The L2 speakers did not show any effects of reading-span or attachment preference in their off-line or on-line decisions in English.

Hopp (2014) investigated the role of WM capacity and lexical automaticity in L2 processing of RC attachment ambiguities by German speakers of L2 English. The results showed that when the RCs required local dependencies reducing the load on memory, the L2 speakers, similar to native speakers, had a low-attachment preference in their on-line decisions. Whereas off-line decisions were affected by WM capacity, lexical automaticity influenced on-line decisions. Hopp concluded that structural processing in the L2 was modulated by WM and lexical automaticity, presumably because L2 processing consumed more resources than native language processing, but not by L2 speakers' inability to compute complex syntactic structure.

Witzel, Witzel, and Nicol (2012) investigated RC attachment in English by L1-Chinese L2-English speakers. The L2 speakers showed a high-attachment preference. Although not necessarily native-like, this was taken as evidence for syntactically-detailed computations. Witzel et al. argue that L1 and L2 speakers could weight the recency and predicate proximity preferences (Gibson, Pearlmuter, Canseco-Gonzalez & Hickok, 1996) differently. Influenced by their L1, the L2 group probably weighted the predicate proximity more heavily.

Pan, Schimke and Felser (2015) investigated how referential context affected RC attachment ambiguity resolution for German and Chinese speakers of English. Both native and non-native speakers' off-line comprehension was influenced by referential context, with an overall NP2 preference by native speakers and NP1 preference by L2 speakers. The L2 speakers' self-paced reading was affected by contextual information; native speakers' on-line decisions were not affected by contextual or syntactic biases. Pan et al. argue that the L2 group's off-line high-attachment preference may be due to the saliency of the NP1 as the head of the complex NP1-*of*-NP2 structure. The L2 on-line data was taken to suggest that L2 sentence processing was influenced by discourse-level information.

As Cunnings (2017) argue, the findings in the studies above –no attachment preference (Felser et al., 2003b; Fernández,

2002; Omaki, 2005; Papadopoulou & Clahsen, 2003) or non-native-like preferences (Dussias, 2003; Pan et al., 2015; Witzel et al., 2012) – may indicate “variable attachment preference” similar to those observed in the L1. But alongside the effects such as discourse context (Pan et al., 2015) or lexical automaticity (Hopp, 2014), the contributing factors to this observation remain to be examined. The present study attempts to fill in this gap with data from a relatively under-studied L1 background, Turkish. The aim is to understand which factor(s) – L1-based parsing, differential weighing of (non)syntactic cues in the L2 or individual differences – explain L2 parsing decisions.

### The present study

There were three participant groups in the study: an L1-Turkish L2-English group with advanced L2 proficiency, a Turkish native-speaker group, and an English native-speaker group. The L2 group took part in all the experiments/tasks in English. Each group participated in a word-by-word self-paced reading experiment to investigate immediate, i.e., real-time, parsing decisions (Just, Carpenter & Woolley, 1982), a read-aloud task to examine their prosodic phrasing and how that affects whole-sentence reading behavior (Webman-Shafran & Fodor, 2016), and a pen-and-paper questionnaire to test sentence-final interpretations. The experimental sentences included a syntactically high or low NP as RC attachment site and manipulated the animacy of the NPs to examine the use of syntactic and semantic information in early and later decisions. The participants also took part in an operational span (OSpan) task to test WM capacity and its effects on these decisions.

An OSpan task was employed due to its reliability in measuring L2 WM capacity (Sanchez, Wiley, Miura, Colflesh, Ricks, Jensen & Conway, 2010). Reading span tasks, as in previous L2 studies on RC attachment ambiguity (Hopp, 2014; Omaki, 2005), are considered by Sanchez et al. to be unreliable in predicting WM capacity due to “the linguistic nature of the processing task which prohibits an accurate assessment of the WMC construct” (2010, p. 492). Turner and Engle (1989) had indeed “created the OSpan task specifically to be less sensitive to influences of reading or language ability.” (Sanchez et al., 2010, p. 492). In Sanchez et al.’s study reading span scores correlated with WM capacity only for the native speaker group, but OSpan scores correlated with it for both native and non-native speakers. The present study employed an automated version of the task (Unsworth, Heitz, Schrock & Engle, 2005). The automated OSpan (AOSPAN) task is computerized and operated by the participants and involves a series of math operations and letter recall (see Unsworth et al., 2005 for details). The instructions were translated to Turkish for presentation to the Turkish participants.

The tasks were administered in the following order: the self-paced reading experiment, the AOSPAN task, the read-aloud production task, and the pen-and-paper questionnaire, with breaks in between them. They are reported separately for Turkish and English.

## Turkish Experiments

### Participants

Seventy-two native speakers of Turkish (34 females, mean age = 25.6) participated in the Turkish experiments and the AOSPAN task. The participants were all born in Turkey and were residing in Turkey at the time of data collection. They received course

credit or small gifts for their participation. All participants reported (corrected-to-)normal vision.

### Turkish Experiment 1: self-paced reading

#### Materials

There were four conditions manipulating the semantic plausibility, i.e., the animacy of NPs in the complex subject (animate, inanimate) and the syntactic structure, i.e., the forced-attachment site (high, low NP). The examples in (1a-b) illustrate the conditions. (The full list of materials is provided in Appendix S1.)

- (1) a. Animate, high/low:  
 [RC Kaza-da yara-lan-an] [N<sub>Plow</sub> araba-nın]  
 Accident-LOC injury-DER-PART car-GEN  
 [N<sub>Phigh</sub> şoför-ü] /  
 driver-3SG.POSS  
 [N<sub>Plow</sub> şoför-ün] [N<sub>Phigh</sub> araba-sı] garaj-da  
 driver-GEN car-3SG.POSS garage-LOC  
 dur-uyor.  
 stand-3SG.PROG  
*The driver of the car / The car of the driver that was injured in the accident is in the garage.*
- b. Inanimate, high/low:  
 [RC Tamir ed-il-en] [N<sub>Plow</sub> şoför-ün]  
 Repair make-PASS-PART driver-GEN  
 [N<sub>Phigh</sub> arabası] /  
 car-3SG.POSS  
 [N<sub>Plow</sub> araba-nın] [N<sub>Phigh</sub> şoförü] garaj-da  
 car-GEN driver-3SG.POSS garage-LOC  
 dur-uyor.  
 stand-3SG.PROG  
*The car of the driver/ The driver of the car that was repaired is in the garage.*

Turkish is a head-final language and RCs precede the NPs that they modify<sup>1</sup>. In the examples in (1a-b), the RC semantically favors either an animate NP, e.g., *driver* as in (1a), or an inanimate NP, e.g., *car* as in (1b) which can be syntactically high (NP2) or low (NP1). The predicate was active in all the sentences and semantically neutral for either ((in)animate) subject. The RC verb’s voice (active/passive) was held constant across the four conditions.

A universal locality principle (late closure or recency) predicts that attaching the RC to the more recent or syntactically lower NP would be more cost-efficient than attaching it to the more distant or syntactically higher NP (Frazier & Fodor, 1978; Gibson et al., 1996; Kimball, 1973). But it is possible that the participants may also use the animacy information associated with the NPs in these decisions, favoring high attachment when the NP is animate (Acuña-Fariña, Fraga, García-Orza & Piñeiro, 2009; Desmet, De Baecke, Drieghe, Brysbaert & Vonk, 2006).

WM capacity may also affect on-line decisions as high WM capacity may lead to low attachment due to longer phrases and no (implicit) prosodic boundary between the RC and the complex NP (Swets et al., 2007; but cf. Traxler, 2007 for prediction of high

<sup>1</sup>Processing this construction may require an active gap-filling strategy in Turkish, but the discussions will focus on RC attachment/modification for parallelism across Turkish and English experiments. The ramifications of potential use of different strategies across the two languages is postponed to General Discussion.

attachment for high WM capacity). High WM capacity may also increase sensitivity to semantic information (King & Just, 1991; Pearlmutter & MacDonald, 1995; Traxler, Williams, Blozis & Morris, 2005) resulting in increased high attachments (Traxler, 2007) when the complex NP head is animate (but cf. Clifton, Traxler, Mohamed, Williams, Morris & Rayner, 2003; Traxler et al., 2005).

There were 32 sentences each in the four conditions above, with a total of 128 experimental sentences. Sentence length was kept constant at six words. The experimental sentences were distributed across four lists counterbalancing for animacy (animate, inanimate) and attachment site (high, low). In each list there were 32 experimental sentences, 64 filler sentences, six practice and six implicit warm-up sentences, totalling up to 108 sentences.

### Procedure

The experiment was conducted using DMDX, version 4.0.4.8 (Forster & Forster, 2003). The participants read the sentences word-by-word through key-press. A *yes/no* comprehension question followed each sentence. The task lasted 20–25 minutes.

### Data analysis

All the participants had  $\geq 95\%$  accuracy. The data were cleaned from incorrect responses and their response times, too quick key presses (< 200 ms.) and time-outs (> 2200 ms). This amounted to 5% of the data.

In Turkish RCs precede the complex genitive NP (RC-NP1-NP2) as in (1). Thus, similar to Japanese, until NP2 the parser does not have evidence that the structure is ambiguous and may behave non-deterministic (Kamide & Mitchell, 1997) or expectation-based (Levy, 2008). The critical regions were determined with these considerations in mind. That is, for sentences such as (1) if low attachment is expected, for high attachment resolutions surprisal will occur at the low NP (NP1). If high attachment is expected, for low attachment resolutions, surprisal will occur at the high NP (NP2). Thus, for conditions where high attachment was forced, the low NP (NP1, the third word, *arabanın* in (1a) and *şoförün* in (1b)) was taken as the critical region; the high NP (NP2, the fourth word (*şoförü/arabası*) was the spillover region. For conditions forcing low attachment, the critical region was the high NP (NP2, the fourth word, *arabası* in (1a) *şoförü* in (1b)); the fifth word (*garajda*) was the spillover region. The RTs at these regions were analyzed.

The data were analyzed using the R statistical computing software, version 2.15.2 (R Core Team, 2012). Initial normality inspections showed deviation from normality for both the critical ( $W = .95$ ,  $p < .001$ ,  $D = .08$ ,  $p < .001$ ) and the spillover region ( $W = .97$ ,  $p < .001$ ,  $D = .08$ ,  $p < .001$ ). Thus, the RT data were log-transformed (Baayen & Milin, 2010; Ratcliff, 1993) and any outliers that were above/below  $\pm 1.5$  x interquartile range were excluded from the data (5.8% of the critical region data and 7.5% of the spillover region data).

For each region, the RTs were entered into a mixed-effects linear regression model. Animacy (animate, inanimate) and attachment (high, low attachment), and WM capacity were fit as fixed factors; subjects and items were random factors. Here and in all the other experiments the levels of the independent variables were coded as 0 and 1 (Cohen & Cohen, 1983; Fox, 1997) and the  $p$ -values reported for the predictors in linear models were calculated using Satterthwaite approximations for degrees of freedom (Luke, 2017). The analyses controlled for faster responses with experiment progress ( $\beta = -.3$ ,  $SE = .03$ ,  $t = -11.67$ ,  $p < .001$  for

the critical region,  $\beta = -.25$ ,  $SE = .02$ ,  $t = -10.2$ ,  $p < .001$  for the spillover region) and for word frequency in the spillover region,  $\beta = -46.05$ ,  $SE = 4.40$ ,  $t = -10.16$ ,  $p < .001$ .

### Results

Figure 1 shows the data for both regions after removing outliers.

Neither animacy nor attachment predicted the RTs for the critical region:  $\beta = 7.63$ ,  $SE = 8.41$ ,  $t = .92$ ,  $p = .36$  for animacy,  $\beta = 15.18$ ,  $SE = 8.3$ ,  $t = 1.86$ ,  $p = .06$  for attachment. The two predictors did not interact, either,  $\chi^2(1) = 2.32$ ,  $p = .13$ . As WM capacity increased the RTs decreased,  $\beta = -5.77$ ,  $SE = 1.69$ ,  $t = -3.4$ ,  $p < .005$  but WM capacity did not interact with attachment,  $\chi^2(1) = 1.57$ ,  $p = .21$ .

For the spillover region, the RTs were faster for low attachment conditions than high attachment conditions,  $\beta = -52.18$ ,  $SE = 10.27$ ,  $t = -5.78$ ,  $p < .001$ . There was no effect of animacy,  $\beta = 3.43$ ,  $SE = 8.26$ ,  $t = .42$ ,  $p = .68$ , nor an interaction involving animacy and attachment  $\chi^2(1) = .09$ ,  $p = .75$ . As WM capacity increased the RTs decreased,  $\beta = -3.33$ ,  $SE = 1.31$ ,  $t = -2.55$ ,  $p < .05$ , but WM capacity did not interact with attachment,  $\chi^2(1) = .07$ ,  $p = .79$ .

During model criticism (for the model with attachment), data points with standardized residuals below/above 2.5 standard deviations were excluded from the analyses as well as overly influential subjects ( $N = 4$  for DR,  $N = 3$  for SR) and individual data points ( $N = 2$  for SR) (Nieuwenhuis, te Grotenhuis & Pelzer, 2012). There were no influential items.

The comprehension questions asked for the participants' final interpretations of the sentences. The data for those answers are presented in Table 1.

The > 93% accuracy shows that the participants recovered from the garden-paths and revised their initial attachment preferences for correct interpretations.

### Turkish Experiment 2: production

#### Materials

Sentences in Experiment 2 were globally ambiguous as the subject NPs were either both animate or inanimate as in (2).

- (2) a. Animate: [RC Kaza-da yara-lan-an] [NP<sub>low</sub> şoför-ün]  
 Accident-LOC injury-DER-PART driver-GEN  
 [NP<sub>high</sub> kardeş-i]  
 brother-3SG.POSS  
 garaj-da dur-uyor.  
 garage-LOC stand-3SG.PROG  
*The brother of the driver that was injured in the accident is in the garage.*
- b. Inanimate: [RC Tamir ed-il-en] [NP<sub>low</sub> araba-nın]  
 Repair make-PASS-PART car-GEN  
 [NP<sub>high</sub> kapı-sı]  
 driver-3SG.POSS  
 garaj-da dur-uyor.  
 garage-LOC stand-3SG.PROG  
*The door of the car that was repaired is in the garage.*

There were 32 experimental item sets, each in two versions, distributed across two lists, counterbalancing for animacy. In each list there were 32 experimental, 64 filler, six practice and six implicit warm-up sentences (108 sentences in total).



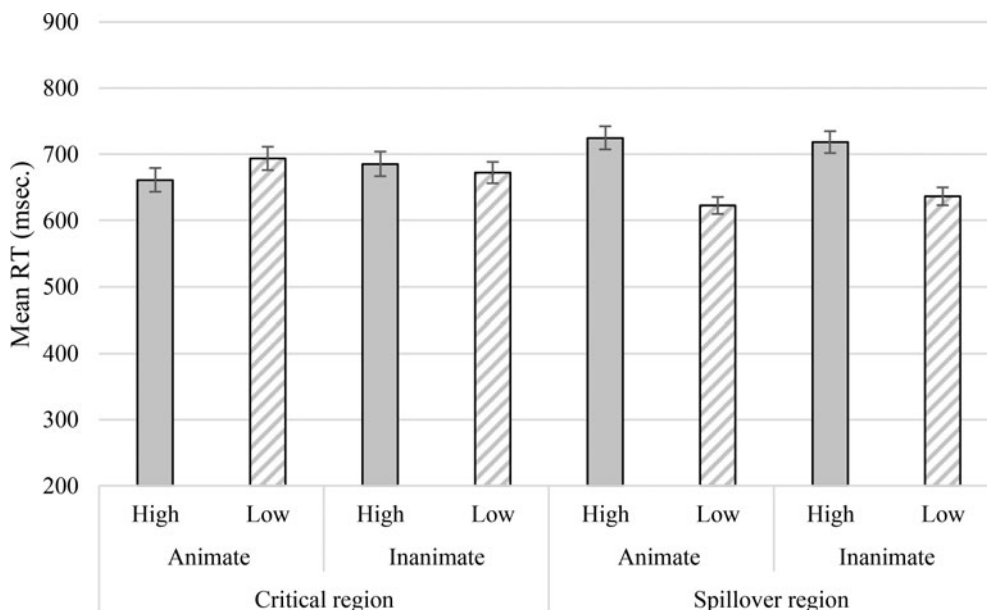


Fig. 1. Turkish Experiment 1, self-paced reading: Mean RTs, with standard errors, for the critical region and spillover region.

Table 1. Turkish L1 group’s percent accuracy, by animacy and forced attachment site, in the self-paced reading task.

	Animate	Inanimate
High	95	95
Low	98	93

Procedure

The participants saw the sentences on a computer screen as a whole. They were asked to read the sentences aloud as soon as the sentences appeared on the screen. They moved from one sentence to the next with a key-press. There were no comprehension questions. The task lasted 8–10 minutes.

Data analysis

The production data were examined for a prosodic boundary after the RC and after NP1 via ear judgments and acoustic analyses. Although the acoustic analyses revealed a pattern consistent with the ear judgment data, the statistical comparisons revealed none to marginal significance presumably due to variance in the acoustic data. Hence the report here is restricted to the ear judgment analyses.

The production data were listened to, independently, by two judges trained in Turkish linguistics and prosody, for the location of a prosodic boundary. The judges were naïve to the purposes of the experiment and were trained for 192 example sentences (~9% of the data) for phonological phrase (PPh) and intonational phrase (IPh) structures in Turkish (Kan, 2009). They then coded their judgements as “break after RC”, indicating high attachment, and “break after NP1” or “no break”, indicating low attachment (see Jun, 2003 for a review). The break could be a PPh boundary, marked with a delimited phrase accent (H-/L-) or an IPh boundary, described as a boundary tone (H%/L%), followed by an optional pause (Beckman & Elam, 1997; Kan, 2009; Kjølgaard & Speer, 1999). The few cases with a break after both the RC and NP1 were coded for the stronger (i.e., IPh)

Table 2. Turkish speakers’ offline RC attachment behavior, as indicated by produced prosody in the read-aloud task and attachment preference in the pen-and-paper questionnaire.

	Percent low-attachment prosody (SE)	Percent low-attachment preference (SE)
Animate	53 (7)	63 (10)
Inanimate	68 (5)	76 (5)

boundary. The inter-rater reliability between the two judges was strong,  $r = .82$ .

The data were analyzed using mixed-effects logistic regression models. Animacy (animate, inanimate) and WM capacity were fit as fixed factors; subjects and items were random factors.

Results

Table 2 shows percent low-attachment prosody with standard errors.

The participants were overall more likely to insert low-attachment prosody ( $M = 61\%$ ) than high-attachment prosody. The likelihood to insert low-attachment prosody increased when the NPs were inanimate, odds ratio:  $\beta = .79$ ,  $SE = .09$ ,  $z = 8.15$ ,  $p < .001$ . As WM capacity increased, the participants were more likely to insert low-attachment prosody, odds ratio:  $\beta = .01$ ,  $SE = .004$ ,  $z = -2.35$ ,  $p < .05$ . There was no interaction between WM and animacy,  $\chi^2(1) = 1.77$ ,  $p = .18$ .

Turkish Experiment 3: off-line reading

Materials

Experiment 3 employed the same sentences as in Experiment 2, as shown in (2) above, followed with comprehension questions such as *Who was injured in the accident? / What was repaired?* The answers to the questions counterbalanced the high and low NPs in the (a) and (b) options.

The number of experimental and filler items were the same as in Experiment 2. There were no practice or implicit warm-up sentences.

### Procedure

The sentences and questions were presented on paper. The task was to indicate which option, (a) or (b), answered the question about the sentence.

### Data analysis

The preference data were analyzed via mixed-effects logistic regression models. Animacy (animate and inanimate) and working memory capacity were fit as fixed factors; subjects and items were random factors.

### Results

The participants' percent low-attachment preference is shown in Table 2.

The participants were overall more likely to attach the RC to the low NP ( $M = 70\%$ ) than the high NP. This preference increased when the NPs were inanimate, odds ratio:  $\beta = .86$ ,  $SE = .11$ ,  $z = 7.94$ ,  $p < .001$ . WM capacity did not predict RC attachment preferences, odds ratio:  $\beta = .01$ ,  $SE = .008$ ,  $z = -1.34$ ,  $p = .18$ , nor did it interact with animacy,  $\chi^2(1) = 1.77$ ,  $p = .18$ .

### Discussion of findings in the Turkish experiments

The results of the on-line experiment showed that Turkish speakers' initial parsing decisions were determined by syntactic biases, late closure/recency (Frazier, 1978; Gibson et al., 1996) or an active gap-filling strategy (Frazier & Clifton, 1989) but their end-of-sentence decisions (viz. prosodic phrasing and pen-and-paper preferences) were also influenced by semantic plausibility (i.e., the animacy of the NPs). As WM capacity increased the likelihood to produce low-attachment prosody also increased (Swets et al., 2007) but it did not affect on-line decisions.

The results overall support the argument for primacy of syntax which maintains that computation of syntactic structure precedes integration of semantic information and the two processes interact only at later stages (Frazier & Fodor, 1978; Friederici, 2002; Sturt, 2003). The results also support heavier weighting of syntactic cues in the L1 (Parker, Shvartsman & Van Dyke, 2017; Van Dyke & McElree, 2011). WM capacity effects on RC attachment preferences through prosodic phrasing are in line with Swets et al.'s (2007) predictions. The null WM effect in on-line decisions could be due to the stimulus being presented word-by-word, disrupting the natural prosody that could be projected by the participants. An eye-tracking experiment where sentential stimulus is presented as a whole could be more informative on that.

## English Experiments

### Participants

Fifty-one Turkish speakers of English with advanced English proficiency and 40 English native-speakers participated in the English experiments and AOSPAN task. All the participants, except one (whose data were excluded from the analyses), lived in an English-speaking country (three resided in Toronto, 47 resided

in New York City) at the time of data collection. Table 3 summarizes the participants' (language) demographics.

The participants received 10\$ for their participation. They all reported (corrected-to-) normal vision.

### English Experiment 1: self-paced reading

#### Materials

There were four conditions manipulating the animacy of NPs in the complex subject (animate vs. inanimate) and the forced attachment site (syntactically high vs. low NP) as in (3a-b). (The full list of materials is provided in Appendix S1.)

- (3) a. Animate, high/low:  
 $[_{\text{NP}^{\text{high}}}$  The driver] of  $[_{\text{NP}^{\text{low}}}$  the car] /  $[_{\text{NP}^{\text{high}}}$  The car] of  $[_{\text{NP}^{\text{low}}}$  the driver]  
 $[_{\text{RC}}$  that was injured] is in the garage.  
 b. Inanimate, high/low:  
 $[_{\text{NP}^{\text{high}}}$  The car] of  $[_{\text{NP}^{\text{low}}}$  the driver] /  $[_{\text{NP}^{\text{high}}}$  The driver] of  $[_{\text{NP}^{\text{low}}}$  the car]  
 $[_{\text{RC}}$  that was repaired ] is in the garage.

The RC semantically favored an animate or inanimate antecedent which was syntactically high (NP1) or low (NP2). The disambiguating information at the RC was limited to one word (a verb or an NP; e.g., *injured/repared* in (3)) to keep its length constant. The disambiguating words were matched in their log10 frequency: average frequency in COCA corpus (Davies, 2008) for animate = 1.47, for inanimate = 1.24.

The predictions for the L1 speakers are the same as those presented for Turkish speakers. The predictions for the L2 speakers are as follows: the L2 speakers may not show any attachment preferences (Clahsen & Felser, 2006a) or they can be sensitive only to semantic information (Clahsen & Felser, 2018) which can inform their syntactic parsing (Cunnings, 2017). If the latter is the case, then high-attaching animacy-forced and low-attaching inanimacy-forced conditions could be processed faster than low-attaching animacy-forced and high-attaching inanimacy-forced conditions. Such data pattern would indicate that the L2 speakers weight semantic cues more heavily than syntactic cues but they also conduct full syntactic parses which are modulated by semantic information. If WM capacity influences such decisions, L2 speakers' data may pattern with low-capacity L1 speakers (Harrington & Sawyer, 1992; Hopp, 2014).

The number of conditions, reading lists, experimental, filler, practice and implicit warm-up sentences in each list was the same as in Turkish self-paced reading experiment. The filler sentences included subject/object relative clauses or subject-verb number (dis)agreement. English experimental sentence length varied from 10–12 words.

### Procedure

The procedure was the same as in Turkish self-paced reading task.

### Data analysis

The participants had  $\geq 95\%$  accuracy. The data were cleaned from incorrect responses and RTs for incorrectly-responded items. This corresponded to 1.5% of the data.

The RTs for the disambiguating region (the 8<sup>th</sup> word, *injured/repared* in (3)) and the spillover region (the 9<sup>th</sup> word, *is* in (3)) were first inspected for normality, separately for the L1 and the L2 group as L2 readers can be slower than L1 readers (e.g.,

**Table 3.** (Language) demographics for English as L1 and L2 groups. POB: place of birth, AOE: age of exposure to English (in years), AOA: age of arrival to an English-speaking country (in years), LOS: length of stay in an English-speaking country (in years), SRP: self-reported English proficiency.

		POB	Mean age	Gender	Highest degree				
<i>Demographics</i>	<i>L1</i>	USA ( <i>N</i> = 38) Canada ( <i>N</i> = 1) S. Korea ( <i>N</i> = 1)	31.5	<i>N</i> of females = 15	High school ( <i>N</i> = 5) College ( <i>N</i> = 17) Master's ( <i>N</i> = 16) PhD ( <i>N</i> = 2)				
	<i>L2</i>	Turkey ( <i>N</i> = 50)	30	<i>N</i> of females = 20	College ( <i>N</i> = 26) Master's ( <i>N</i> = 16) PhD ( <i>N</i> = 8)				
		Mean AOE	Mean AOA	Mean LOS	School Language	SPR	Score in TOEFL	Weekly Turkish use (in hours)	Weekly English use (in hours)
<i>Language demographics for L2</i>		11.4	24.7	5.3	Elementary School: Turkish ( <i>N</i> = 50) Secondary School: Turkish ( <i>N</i> = 33) English ( <i>N</i> = 15) German ( <i>N</i> = 1) French ( <i>N</i> = 1) College: Turkish ( <i>N</i> = 5) English ( <i>N</i> = 45) Graduate School: Turkish ( <i>N</i> = 2) French ( <i>N</i> = 1) English ( <i>N</i> = 41)	Advanced ( <i>N</i> = 50)	IBT > 90 CBT > 225 PBT > 554	Home <i>M</i> = 19.2 Social <i>M</i> = 6.5 Work <i>M</i> = 0.69	Home <i>M</i> = 3.45 Social <i>M</i> = 6.4 Work <i>M</i> = 24.11

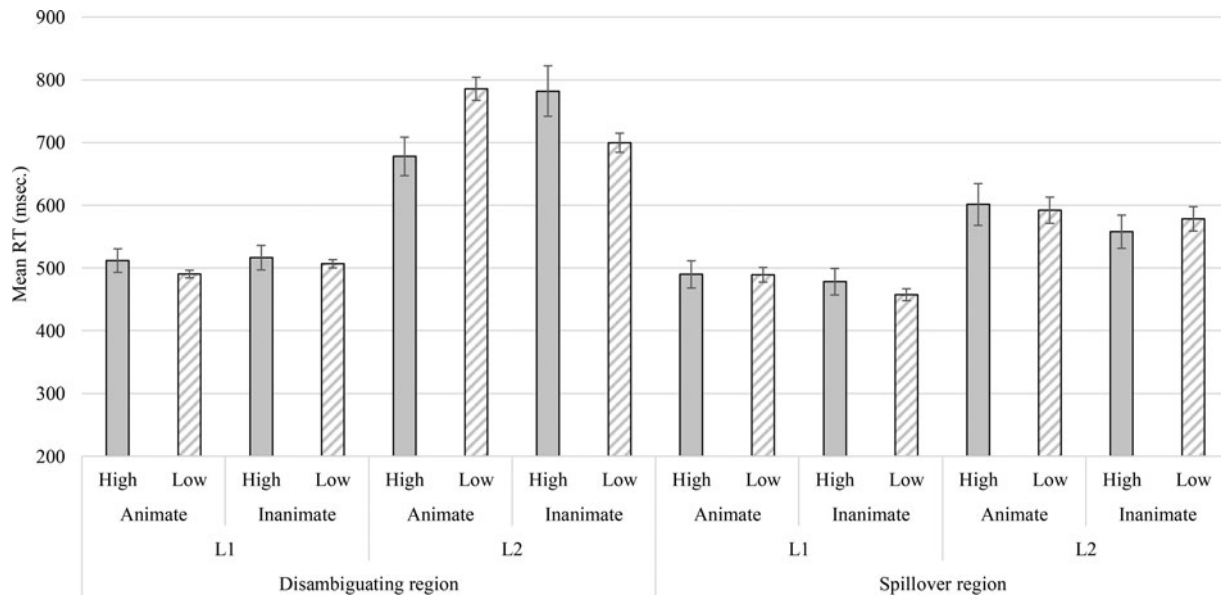


Fig. 2. English Experiment 1, self-paced reading: L1 and L2 speakers' mean RTs, with standard errors, for the disambiguating region and spillover region.

Dinçtopal-Deniz, 2010; Keating, 2009). The RTs deviated from normality for both the disambiguating (L1 group:  $W = .58$ ,  $p < .001$ ,  $D = .21$ ,  $p < .001$ ; L2 group:  $W = .54$ ,  $p < .001$ ,  $D = .26$ ,  $p < .001$ ) and the spillover region (L1 group:  $W = .54$ ,  $p < .001$ ,  $D = .23$ ,  $p < .001$ ; L2 group:  $W = .92$ ,  $p < .001$ ,  $D = .13$ ,  $p < .001$ ). They were hence log-transformed and any outliers that were above/below  $\pm 1.5$  x the interquartile range were excluded (3% of disambiguating and 6% of the spillover data for L1, and 1% of disambiguating and 6% of the spillover data for L2).

The main analyses were conducted on the L1 and L2 data together. The data were split by group for further analyses where necessary. For each region, the RTs were entered into a mixed-effects model. All the R packages,  $p$  value calculation procedures and model building steps were the same as in Turkish Experiment 1. Group (L1, L2), animacy (animate, inanimate), attachment (high, low), and working memory capacity were fixed factors; subjects and items were random factors. The analyses controlled for faster responses with progress ( $\beta = -.27$ ,  $SE = .03$ ,  $t = -10.06$ ,  $p < .001$  for the disambiguating region,  $\beta = -.09$ ,  $SE = .01$ ,  $t = -7.05$ ,  $p < .001$  for the spillover region).

## Results

Figure 2 shows the disambiguating and spillover region data, after removing outliers.

**Disambiguating region:** the L2 group read the disambiguating region slower than the L1 group,  $\beta = -.236$ ,  $SE = 84.72$ ,  $t = 3.54$ ,  $p < .001$ . Attachment, animacy or WM did not explain the data,  $t$ 's  $< 1.12$  but group, attachment and animacy predictors interacted,  $\chi^2(4) = 10.93$ ,  $p < .05$ . The analyses that followed were conducted on L1 and L2 data separately to better understand the nature of the interaction.

The analyses on L1 data showed that disambiguation towards low NP shortened the RTs in the disambiguating region,  $\beta = -25.43$ ,  $SE = 9.24$ ,  $t = -2.75$ ,  $p < .005$ . There was no effect of animacy ( $\beta = 11.08$ ,  $SE = 10.65$ ,  $t = 1.06$ ,  $p = .15$ ) nor did it contribute to a complex model with attachment ( $\chi^2(1) = 1.16$ ,  $p = .28$ ). As WM capacity increased the RTs decreased,

$\beta = -6.38$ ,  $SE = 2.74$ ,  $t = -2.33$ ,  $p < .05$ . WM did not interact with attachment, though,  $\chi^2(1) = 2.8$ ,  $p = .09$ .

The analyses for the L2 group indicated that a model with attachment and animacy interaction predicted the RTs at the disambiguating region better than a simpler model with animacy ( $\chi^2(2) = 7.19$ ,  $p < .05$ ) or attachment ( $\chi^2(2) = 7.15$ ,  $p < .05$ ). Planned comparisons showed that in animacy-forced conditions, disambiguation towards low attachment increased the RTs ( $\beta = 107.74$ ,  $SE = 36.76$ ,  $t = 3.23$ ,  $p < .005$ ) whereas in inanimacy-forced conditions, it decreased the RTs ( $\beta = -82.45$ ,  $SE = 30.43$ ,  $t = 2.62$ ,  $p < .01$ ). WM did not contribute to the model with attachment,  $\chi^2(1) = 0.74$ ,  $p = .39$ .

**Spillover region:** analyses on the spillover region showed an effect of group, where again the L2 speakers read the region slower than the L1 speakers,  $\beta = 101.04$ ,  $SE = 36.35$ ,  $t = 3.15$ ,  $p < .005$ . There was also an effect of animacy, where the conditions attaching the RC to inanimate NPs were read faster than those that attached the RC to animate NPs,  $\beta = -21.49$ ,  $SE = 8.22$ ,  $t = 2.58$ ,  $p < .01$ . There was no effect of attachment,  $\beta = -1.26$ ,  $SE = 8.39$ ,  $t = -.15$ ,  $p = .88$ . The models with an interaction of group and attachment, and of group and animacy explained the data better than a simple model with attachment ( $\chi^2(2) = 10.13$ ,  $p < .01$ ) or animacy ( $\chi^2(2) = 9.7$ ,  $p = .01$ ). Hence the data were analyzed separately for the L1 and L2 speakers.

The data for the L1 group did not show any effects of attachment or animacy,  $t$ 's  $< 1.73$ . The model with animacy and attachment together showed only marginal improvement,  $\chi^2(1) = 2.99$ ,  $p = .08$ . As WM capacity increased, the RTs decreased  $\beta = -2.65$ ,  $SE = 1.29$ ,  $t = -2.05$ ,  $p < .05$  but it did not interact with attachment,  $\chi^2(1) = .02$ ,  $p = .89$ .

The data for the L2 group showed an effect of animacy,  $\beta = -27.55$ ,  $SE = 13.76$ ,  $t = 1.98$ ,  $p < .05$ ; the RTs decreased for inanimate NPs. Neither attachment nor WM explained the data,  $t$ 's  $< 1.12$ . Animacy and attachment did not interact, either,  $\chi^2(1) = 1.12$ ,  $p = .29$ .

The data for final interpretations of the sentences are presented in Table 4.



**Table 4.** L1 and L2 group's percent accuracy, by animacy and forced attachment site, in the self-paced reading task.

		Animate	Inanimate
L1	High	77	82
	Low	86	93
L2	High	78	80
	Low	87	85

The percent accuracy on comprehension questions showed that both groups recovered from garden-paths (> 77% accuracy). For the L1 group, accuracy was the lowest for high-attaching animate conditions and the highest for low-attaching inanimate conditions. The L2 data showed a similar pattern.

**English Experiment 2: production**

**Materials**

Sentences in the production experiment were globally ambiguous as the complex NPs were either both animate as in (4a) or both inanimate as in (4b).

- (4) a. Animate: The brother of the driver that was injured is in the garage.
- b. Inanimate: The door of the car that was repaired is in the garage.

L2 speakers have been reported to use syntactic cues in later decisions (Felser, Sato & Bertenshaw, 2009). If so, similar to native speakers, both syntactic and semantic information can influence their produced prosody. But if WM also affects these decisions and if they behave similar to low WM capacity readers, they may insert a prosodic boundary before the RC leading to increased high-attachment prosody.

The number of conditions, lists, experimental, practice and warm-up sentences were the same as in the Turkish production experiment. There were 32 fillers in the English production experiment. To ensure comprehension the fillers had a missing copula/auxiliary to be filled in while reading. There were 76 sentences in total in each list.

**Procedure**

The procedure was the same as in the Turkish production task.

**Data analysis**

The production data were examined for a prosodic boundary after NP1 (high NP) and NP2 (low NP). The participants were >97% correct in fillers ( $M_{L1} = 95\%$ ,  $M_{L2} = 99\%$ ). The data were examined both via ear judgments and acoustically but the ear judgment analyses are reported here for reasons indicated in the Turkish production experiment. (The acoustic data patterned similar to the ear judgment data.)

The judges were the same as those in the Turkish production experiment. They were trained on 419 example sentences (~15% of the data) for PPh and IPh structures. They then coded their judgements as “break after NP2”, indicating high attachment or “break after NP1” or “no break”, indicating low attachment. The agreement between the judges was high,  $r = .81$  for the L1 data,  $r = .86$  for the L2 data.

**Table 5.** L1 and L2 group's offline RC attachment behavior, as indicated by produced prosody in the read-aloud task and attachment preference in the pen-and-paper questionnaire.

		Percent low-attachment prosody (SE)	Percent low-attachment preference (SE)
L1	Animate	51 (2)	72 (1.8)
	Inanimate	64 (1.9)	69 (1.8)
L2	Animate	34 (1.7)	50 (1.8)
	Inanimate	48 (1.8)	52 (1.8)

The data were analyzed using mixed-effects logistic regression models. Group (L1 and L2), animacy (animate, inanimate) and working memory capacity were fit as fixed factors; subjects and items were random factors.

**Results**

Table 5 shows percent low-attachment prosody, with standard errors.

The analyses showed that the L2 speakers were less likely than the L1 speakers to produce low-attachment prosody, odds ratio:  $\beta = -.84$ ,  $SE = .19$ ,  $z = -4.33$ ,  $p < .001$  and the likelihood to insert low-attachment prosody increased with inanimate NPs, odds ratio:  $\beta = .69$ ,  $SE = .09$ ,  $z = 8.1$ ,  $p < .001$ . The model with animacy and group explained the data better than the simpler model with animacy,  $\chi^2(1) = 17.11$ ,  $p < .001$ , or group,  $\chi^2(1) = 66.43$ ,  $p < .001$ . WM did not predict the data,  $\beta = -.002$ ,  $SE = .006$ ,  $z = 0.32$ ,  $p = .75$ ; nor did it interact with any other predictor,  $\chi^2s < .42$ ,  $p's > .52$ . Since the complex model with group and animacy explained the data better, the data were examined separately for the L1 and L2 speakers.

The L1 group inserted low-attachment prosody a little over half the time ( $M = 56\%$ ). This was higher for conditions with inanimate NPs ( $M = 64\%$ ) than animate NPs ( $M = 52\%$ ). Analyses confirmed that the likelihood to insert a low-attachment boundary increased when the NPs were inanimate,  $\beta = .65$ ,  $SE = .13$ ,  $z = 4.99$ ,  $p < .001$ . The model with WM did not explain the data,  $\beta = .0008$ ,  $SE = .0008$ ,  $z = .10$ ,  $p = .92$ ; it did not improve the model with animacy, either,  $\chi^2(1) = .01$ ,  $p = .92$ .

The L2 group was overall less likely to insert low-attachment prosody ( $M = 40\%$ ). This was stronger for conditions with animate NPs ( $M = 33\%$ ) than inanimate NPs ( $M = 48\%$ ). As in the L1 group, the likelihood to insert low-attachment boundary increased when the NPs were inanimate,  $\beta = .74$ ,  $SE = .12$ ,  $z = 6.46$ ,  $p < .001$ . WM did not explain the data,  $\beta = -.003$ ,  $SE = .008$ ,  $z = -.40$ ,  $p = .67$ , nor did it improve the animacy model,  $\chi^2(2) = 2.73$ ,  $p = .26$ .

**English Experiment 3: off-line reading**

**Materials**

The same sentences as in (4) were used but they were followed with questions such as *Who was injured in the accident? / What was repaired?*

The number of conditions, lists, experimental and filler sentences were the same as in the English production experiment. There were no practice or warm-up sentences. The high and low NPs were counterbalanced across (a) and (b) options.

### Procedure

The procedure was the same as in the Turkish off-line task.

### Data analysis

The same data analysis procedures as in Turkish off-line experiment were used. Group (L1, L2), animacy (animate, inanimate), and WM capacity were fit as fixed factors; subjects and items were random factors.

### Results

The participants' percent low-attachment preference is shown in Table 5.

The L1 group preferred to attach the RC low ( $M = 71\%$ ) but the L2 group did not show any preference ( $M = 51\%$  low attachment). The analyses showed that the L2 group was less likely to attach low than the L1 group, odds ratio:  $\beta = 1.03$ ,  $SE = .31$ ,  $z = 3.25$ ,  $p < .005$ . Animacy and WM did not predict attachment preference,  $z$ 's  $< .29$ ,  $p$ 's  $> .78$ . The complex model with group and animacy together did not explain the data better than the simpler model with group,  $\chi^2(1) = .08$ ,  $p = .78$ .

### Discussion of the findings in English experiments

The RT data showed that native speakers of English, similar to native speakers of Turkish, were guided by syntactic biases, i.e., late closure/recency (Frazier, 1978; Gibson et al., 1996); their on-line decisions were not affected by semantic information or WM capacity. There was no syntactic bias in the L2 group's on-line decisions. Their real-time processing of the ambiguity was mainly modulated by semantic information: when the RC modified the animate NP, they preferred to attach high but when it manipulated the inanimate NP, they preferred to attach low.

The production data for both groups showed an effect of semantic plausibility. The L1 group produced low-attachment prosody when the NPs were inanimate; there was no prominent prosodic phrasing when the NPs were animate. The L2 speakers produced high-attachment prosody when the NPs were animate; there was no preferred prosody when the NPs were inanimate. There was no effect of WM capacity for either group. Although semantic information carried by the NPs affected the produced prosody for both groups, the syntactic bias was an additional force for native speakers as they showed no clear prosodic phrasing for animate NPs. The late closure/recency bias probably cancelled out (but not reversed) the animacy effect. For the L2 group, the referentiality of the head (i.e., high) NP appears to be an additional force (Frazier & Clifton, 1996; Gilboy, Sopena, Clifton & Frazier, 1995) as there was an overall high-attachment preference and the inanimacy of the NPs only cancelled it out.

The end-of-sentence questionnaire data replicated the previously reported low-attachment preference for English. The L2 speakers showed no attachment preference in these decisions. There was no effect of animacy or WM capacity for either group.

The data for the native speaker group support the primacy of syntax view (Frazier & Fodor, 1978; Friederici, 2002; Sturt, 2003) as their on-line decisions were influenced by syntactic biases and semantic information affected later decisions as revealed by the prosody data. Semantic information was used in on-line decisions by the L2 speaker group. The animacy of the NP head alone modulated the RTs for the disambiguating region. It is possible to refer to the SSH to explain these results since the SSH predicts for the L2 speakers to "rely more on lexico-semantic cues to

interpretation" (Clahsen & Felser, 2006a, p. 107) but the SSH does not predict the L2 speakers' ability to compute, on-line, RC attachment to complex genitive NPs due to representations that lack syntactic detail. The L2 speakers showed an attachment preference though, either high or low depending on semantic information. Thus, the SSH appears to be inadequate in explaining the findings. The results may be explained better in the context of cue-based memory retrieval in the L2 (Cunnings, 2017), which is discussed in detail below.

### General discussion

Both groups of L1 speakers relied on syntactic information in their on-line decisions. The low-attachment preference by L1 Turkish speakers could be the result of the late closure/recency preference or an active gap-filling strategy because in Turkish the RC precedes the complex genitive NP (RC-NP1-NP2) and postulates a gap that can be filled by either the NP1 (low NP) or the NP2 (high NP). The low NP is the first potential filler for the gap postulated by the RC. The low-attachment preference in Turkish could therefore be an instance of the *Active Filler* strategy (Frazier & Clifton, 1989) by which the parser would try to actively match a filler (gap in Turkish) with a gap (filler in Turkish). Because keeping the filler (or gap) in memory is costly, an interpretation in which the first gap (or filler) could resolve the dependency would be preferable. In Turkish the first NP (low NP) serves as the first legitimate filler for the gap introduced by the RC, presumably leading to the low-attachment preference.

Cunnings (2017) argue that structures that involve backwards-looking (e.g., anaphoric) dependencies restrict WHAT information is to be retrieved once a memory retrieval operation is initiated (e.g., upon encountering the reflexive, the parser must look backwards to find an antecedent) whereas structures that involve forward-looking (e.g., filler-gap) dependencies restrict WHEN a memory retrieval operation can be initiated (e.g., once a filler is encountered, the parser must actively look for a gap). Given Cunnings's (2017) arguments on the WHAT's and WHEN's of cue retrieval, the operations that an English and a Turkish parser processing an RC employs may not be the same. (Also see Levy's (2008) surprisal theory for expectation-based processing in head-final languages.) An English native speaker, having processed the complex NP (subject or object) would, upon encountering the RC, decide where to ATTACH the RC. The Turkish speaker, on the other hand, would postulate a filler-gap dependency and resolve that dependency. This suggests that the processing of English RC attachment ambiguities by L1 Turkish speakers could not be traced to L1-based parsing preferences as the ambiguity would need to be treated as a different structure in the L2 (see Omaki, 2005, p. 91 for a similar discussion).

The L2 group's on-line decisions were informed by semantic information. The L2 speakers preferred to attach high when the RC defined the animate NP but they preferred to attach low when the RC manipulated the inanimate NP. Although the SSH predicts an increased reliance on semantic information in L2 sentence processing, that under SSH the L2 speakers are not expected to show clear attachment preferences rules out a shallow parsing explanation. L2 speakers' responses to the comprehension questions in the self-paced reading task (both in pattern and in accuracy) also indicate that their recovery from potential garden-paths (>78%) was similar to that by L1 speakers (>77%). This further suggests that they did not do more good-enough parsing than the L1 group (cf. Jacob & Felser, 2016; Pozzan & Trueswell, 2016).

As was mentioned in the Introduction, Cummings argues that L2 learners compute fully-specified syntactic parses but they are more susceptible to interference during cue retrieval from memory. Although RC attachment preferences in the L2 are viewed by Cummings as simply variability in choice which is not too different from those in the L1 (cf. Grillo & Costa, 2014, though) and although Cummings does not make specific predictions with respect to weighting of semantic cues such as animacy as opposed to syntactic cues, the present data show that semantic information either outweighs syntactic information or prevails in absence of syntactic biases. But neither interpretation suggests unavailability of complex syntactic representations in the L2; the L2 speakers in the present study did process complex syntax, albeit differently from native speakers. The possibility of heavier weighting of the animacy cue in the L2 will be revisited after discussing the prosody and off-line data.

Both L1 Turkish and L1 English speakers were more likely to produce low-attachment prosody but this tendency was reduced with animate NPs. The L2 speakers were overall more likely to insert high-attachment prosody but this tendency was reduced for inanimate NPs. The null effect of WM capacity in prosodic phrasing in the L1 or L2 data rules out WM capacity-related explanations for prosodic chunking. The general tendency for high-attachment prosody by L2 speakers may be attributed to the referentiality or topicality of the high NP (as it is head of the complex subject NP), and to referentiality being a more heavily-weighted cue in the L2 than a syntactic bias such as late closure.

L1 Turkish and L1 English groups' off-line questionnaire data showed a similar pattern to their production data. There was no reliable attachment preference in the L2 speakers' off-line data; this was unpredicted but can suggest that prosodic phrasing may better indicate sentence-final decisions than pen-and-paper questionnaires, specifically for L2 speakers. It is possible that in responding to untimed questionnaires L2 speakers refer to their prescriptive/declarative knowledge of how such constructions SHOULD be interpreted. Some participants indeed indicated that they referred to prescriptive rules of English in their responses. They had learned that the RCs modify the immediately preceding constituent, which seems to diverge from their prosodic phrasings. This may have resulted in a trade-off in semantics/prosody-informed decisions and prescriptive knowledge, inducing a lack of preference in L2 off-line decisions.

WM capacity overall had a null effect, except for Turkish production data. It is possible to attribute this to the participants' overall high WM scores. Most participants were either college or graduate students or individuals with a college/graduate degree. It is likely that the range of WM scores, especially in the English experiments (20–66, with *mode* = 48), were not sufficient to capture subtle WM effects. Alternatively, “the amount of memory required for incremental parsing is extremely limited, making the size of an individual's working memory inconsequential for interpretation” (Parker et al., 2017, p. 122) as would be assumed in cue-based memory retrieval model (Lewis, Vasishth & Van Dyke, 2006; McElree, Foraker & Dyer, 2003). This is consistent with the overall data which highlights heavier weighting of syntactic cues in L1 (Parker et al., 2017; Van Dyke & McElree, 2011) and semantic cues in L2 (Cummings, 2017).

The results overall showed that the semantic information was influential and interacted with syntactic biases in later decisions for the L1 speakers (production data and pen-and-paper decisions). It was informative in both real-time and final decisions

for the L2 speakers (self-paced reading and production data). The question then is how the lexico-semantic cue – namely, the animacy of the NPs in the subject – has affected the L2 speakers' on-line and off-line judgments (and the L1 speakers' later decisions). One possibility is that the animate NPs are more referential and topic-worthy than inanimate NPs. Animacy has been reported to statistically correlate “with topicality (or referentiality/definiteness) and syntactic weight. Animates, especially humans, are more likely to be a discourse topic and are therefore also more likely to be definite.” (Rosenbach, 2008, p. 156). In Norman genitive constructions, such as in the English experimental sentences in this study, animate NPs are predicted to precede inanimate NPs and are more likely to be heads (Kirby, 1996). In their REFERENTIALITY PRINCIPLE Gilboy et al. (1995) argue that more referential NPs attract RC attachment and the heads of complex genitive NPs are referential as they introduce discourse entities or correspond to already existing entities. Similarly, Bock, Loebell and Morey (1992) argue that there is a statistical relationship between animacy and subject-hood where animates are more natural subjects than inanimates. In Keil's 1979 hierarchy of predicates (cited in Bock et al., 1992), animates are at the bottom and more predicates apply to them than inanimates. Bock et al. maintain that “[t]he abundance of predicates that can be applied to animate concepts argues that human cognition centers on animate beings, with the thoughts we have and the messages we compose reflecting this fact.” (p. 156). Thus, animacy could be a factor determining conceptual accessibility (Bock, 1987; Bock et al., 1992). The referentiality of high NPs was, therefore, likely enhanced by two forces (i) through their being animate and (ii) through their being heads of maximal projections. But this prevailed as a stronger cue in the L2 parsing of the construction than in its L1 parsing and in its absence, e.g., in the context of inanimate NPs, the L2 speakers, just like the L1 speakers, appear to be influenced by locality principles (Frenck-Mestre & Pynte, 1997).

Topic-worthiness/discourse prominence of animate heads would be in line with Cummings' claims on L2 speakers' heavier weighting of discourse-level cues as the animacy of a head NP would increase its discourse prominence and Cummings argues “[f]or L2 speakers, the preference for a subject/topic seems to be a more heavily weighted cue than in L1 speakers” (p. 669). The question then is why L2 speakers weight that cue heavier than the syntactic one. Cummings does not appear to provide an answer to that, neither can the present data.

Witzel et al. (2012) and Omaki (2016) suggest that non-native-like processing may be due to L1 weighting of the cues but the on-line L1 data for Turkish did not reveal a role of animacy. Omaki (2016) also suggests that differential weighting of cues in the L2 may be learned. That is, if ‘discourse-related’ cues have statistically proven to be more successful in reaching meaning, then L2 learners may rely more heavily on those than syntactic cues. Similarly, Grüter et al. (2020) argue that L2 speakers weigh cues differently from L1 speakers to be more efficient in processing “[i]f certain type of information, say lexical semantics, presents a more reliable source of knowledge to base a prediction on than another type of information ... then the optimally adaptive language user will rely more heavily on the former than the latter” (p.232).

It is possible that greater reliance on semantic cues in L2 processing can be “an effect of adaptation to the relative reliability of information, serving to maximize L2 processing proficiency” (Grüter et al., 2020, p. 221). This could be because L2 processing



may be less modular than L1 processing and this may reveal itself in structures such as RC attachment ambiguities as they are, at least those in English, a form of adjunct attachment and do not involve primary relations (the Construal Hypothesis by Frazier & Clifton, 1996). Several studies have shown that children, compared to adults, rely more heavily on syntactic cues than semantic/pragmatic cues (e.g., Felser, Marinis, Clahsen, Felser, Marinis & Clahsen, 2003a; Traxler, 2002; Trueswell, Sekerina, Hill & Logrip, 1999). They appear to develop the ability to integrate semantic information later on, presumably with increased efficiency in lexical retrieval. (It has further been reported that, once meaning for a sentence is reached, its exact syntactic form is forgotten (Sachs, 1967, 1974)). This developmental change in parsing may affect cue-weighting in processing. And second language processing, especially when the L2 is acquired after certain cognitive developments have taken place, may prioritize meaning over form.

**Supplementary Material.** For supplementary material accompanying this paper, visit <https://doi.org/10.1017/S1366728921001140>

**Competing interests.** The author declares none.

**Acknowledgements.** I would like to thank Eva M. Fernández for consultation during the initial stages of this project. I am grateful to Özge Bakay, Didem Bayrak Kurt and Didar Karadağ for their help with data collection and their meticulous annotations and ear-judgments on the speech data.

This work was supported by Boğaziçi University Scientific Research Projects Grant [#13781]. Partial financial support was received from the City University of New York, Graduate Center, Student Research fund.

## References

- Acuña-Fariña C, Fraga I, García-Orza J and Piñeiro A (2009) Animacy in the adjunction of Spanish RCs to complex NPs. *European Journal of Cognitive Psychology* 21(8), 1137–1165. <https://doi.org/10.1080/09541440802622824>
- Baayen RH and Milin P (2010) Analyzing reaction times. *International Journal of Psychological Research* 3(2), 12–28.
- Beckman ME and Elam GA (1997) Guidelines for ToBI labelling (version 3). Retrieved from [http://www.ling.ohio-state.edu/research/phonetics/E\\_ToBI/](http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/)
- Bock K (1987) An effect of the accessibility of word forms on sentence. *Journal of Memory and Language* 137(26), 119–137.
- Bock K, Loebell H and Morey R (1992) From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review* 99(1), 150–171. <https://doi.org/10.1037/0033-295X.99.1.150>
- Clahsen H and Felser C (2006a) Continuity and shallow structures in language processing. *Applied Psycholinguistics* 27(1), 107–126. <https://doi.org/10.1017/S0142716406060206>
- Clahsen H and Felser C (2006b) Grammatical processing in language learners. *Applied Psycholinguistics* 27(1), 3–42. <https://doi.org/10.1017/S0142716406060024>
- Clahsen H and Felser C (2006c) How native-like is non-native language processing? *Trends in Cognitive Sciences* 10(12), 564–570. <https://doi.org/10.1016/j.tics.2006.10.002>
- Clahsen H and Felser C (2018) Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition* 40(3), 693–706. <https://doi.org/10.1017/S0272263117000250>
- Clifton C, Traxler MJ, Mohamed MT, Williams RS, Morris RK and Rayner K (2003) The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language* 49(3), 317–334. [https://doi.org/10.1016/S0749-596X\(03\)00070-6](https://doi.org/10.1016/S0749-596X(03)00070-6)
- Cohen J and Cohen P (1983) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.
- Cunings I (2017) Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition* 20(4), 659–678. <https://doi.org/10.1017/S1366728916000675>
- Davies M (2008) The Corpus of Contemporary American English (COCA). Retrieved from <https://www.english-corpora.org/>
- Desmet T, De Baecke C, Drieghe D, Brysbaert M and Vonk W (2006) Relative clause attachment in Dutch: On-line comprehension corresponds to corpus frequencies when lexical variables are taken into account. *Language and Cognitive Processes* 21(4), 453–485. <https://doi.org/10.1080/01690960400023485>
- Dinçtopal-Deniz N (2010) Relative clause attachment preferences of Turkish L2 speakers of English: Shallow parsing in the L2? In VanPatten B & Jegerski J (eds), *Research in second language processing and parsing*. Amsterdam: John Benjamins, pp. 27–63.
- Dussias PE (2003) Syntactic ambiguity resolution in L2 learners. *Studies in Second Language Acquisition* 25(4), 529–557. <https://doi.org/10.1017/S0272263103000238>
- Felser C, Marinis T, Clahsen H, Felser C, Marinis T and Clahsen H (2003a) Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition* 11(3), 127–163. <https://doi.org/10.1207/s15327817la1103>
- Felser C, Roberts L, Marinis T and Gross R (2003b) The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics* 24(3), 453–489. <https://doi.org/10.1017/S0142716403000237>
- Felser C, Sato M and Bertenshaw N (2009) The on-line application of binding principle A in English as a second language. *Bilingualism* 12(4), 485–502. <https://doi.org/10.1017/S1366728909990228>
- Fernández EM (2002) Relative clause attachment in bilinguals and monolinguals. In Heredia RR & Altarriba J (eds), *Bilingual sentence processing*. Amsterdam, The Netherlands; New York: N-H (North-Holland), pp. 187–215.
- Forster KI and Forster JC (2003) DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers* 35(1), 116–124. <https://doi.org/10.3758/BF03195503>
- Fox J (1997) *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks: Sage Publications.
- Frazier L (1978) *On comprehending sentences: Syntactic parsing strategies*. (Unpublished doctoral dissertation). University of Connecticut, Storrs.
- Frazier L and Clifton C (1989) Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4(2), 93–126. <https://doi.org/10.1080/01690968908406359>
- Frazier L and Clifton C (1996) *Construal*. Cambridge, MA: MIT Press.
- Frazier L and Fodor JD (1978) The sausage machine: A new two-stage parsing model. *Cognition* 6(4), 291–325. [https://doi.org/10.1016/0010-0277\(78\)90002-1](https://doi.org/10.1016/0010-0277(78)90002-1)
- Frenc-Mestre C and Pynte J (1997) Syntactic ambiguity resolution while reading in second and native Languages. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology* 50(1), 119–148. <https://doi.org/10.1080/027249897392251>
- Friederici AD (2002) Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6(2), 78–84. [https://doi.org/10.1016/S1364-6613\(00\)01839-8](https://doi.org/10.1016/S1364-6613(00)01839-8)
- Gibson E, Pearlmutter N, Canseco-Gonzalez E and Hickok G (1996) Recency preference in the human sentence processing mechanism. *Cognition* 59(1), 23–59. [https://doi.org/10.1016/0010-0277\(95\)00687-7](https://doi.org/10.1016/0010-0277(95)00687-7)
- Gilboy E, Sopena J-M, Clifton C and Frazier L (1995) Argument structure and association preferences in Spanish and English complex NPs. *Cognition* 54(2), 131–167. [https://doi.org/10.1016/0010-0277\(94\)00636-y](https://doi.org/10.1016/0010-0277(94)00636-y)
- Grillo N and Costa J (2014) A novel argument for the universality of parsing principles. *Cognition* 133(1), 156–187. <https://doi.org/10.1016/j.cognition.2014.05.019>
- Grüter T, Lau E and Ling W (2020) How classifiers facilitate predictive processing in L1 and L2 Chinese: The role of semantic and grammatical cues. *Language, Cognition and Neuroscience* 35(3), 221–234. <https://doi.org/10.1080/23273798.2019.1648840>
- Harrington M and Sawyer M (1992) L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition* 14(1), 25–38. <https://doi.org/10.1017/S0272263100010457>
- Hopp H (2014) Working memory effects in the L2 processing of ambiguous relative clauses. *Language Acquisition* 21(3), 250–278. <https://doi.org/10.1080/10489223.2014.892943>



- Jacob G and Felsler C (2016) Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology* 69(5), 907–925. <https://doi.org/10.1080/17470218.2014.984231>
- Jun SA (2003) Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research* 32(2), 219–249. <https://doi.org/10.1023/A:1022452408944>
- Just MA, Carpenter PA and Woolley JD (1982) Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111(2), 228–238. <https://doi.org/10.1037/0096-3445.111.2.228>
- Kamide Y and Mitchell DC (1997) Relative clause attachment: Nondeterminism in Japanese parsing. *Journal of Psycholinguistic Research* 26(2), 247–254. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=11303230&site=ehost-live>
- Kan S (2009) *Prosodic domains and syntax prosody mapping in Turkish*. (Unpublished M.A. thesis). Boğaziçi University, Istanbul.
- Keating GD (2009) Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning* 59(3), 503–535. <https://doi.org/10.1111/j.1467-9922.2009.00516.x>
- Kimball J (1973) Seven principles of surface structure parsing in natural language. *Cognition* 2(1), 15–47. [https://doi.org/10.1016/0010-0277\(72\)90028-5](https://doi.org/10.1016/0010-0277(72)90028-5)
- King J and Just MA (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](https://doi.org/10.1016/0749-596x(91)90027-h)
- Kirby S (1996) *Function, Selection and Innateness: The Emergence of Language Universals*. (Unpublished doctoral dissertation). University of Edinburgh. <https://doi.org/10.1525/aa.1964.66.2.02a00060>
- Kırkacı B (2004) The processing of relative clause attachment ambiguities in Turkish. *Turkic Languages* 8, 111–121.
- Kjelgaard MM and Speer SR (1999) Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* 40(2), 153–194. <https://doi.org/10.1006/jmla.1998.2620>
- Levy R (2008) Expectation-based syntactic comprehension. *Cognition* 106(3), 1126–1177. <https://doi.org/10.1016/j.cognition.2007.05.006>
- Lewis RL, Vasishth S and Van Dyke JA (2006) Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences* 10(10), 447–454. <https://doi.org/10.1016/j.tics.2006.08.007>
- Luke SG (2017) Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods* 49(4), 1494–1502. <https://doi.org/10.3758/s13428-016-0809-y>
- McElree B, Foraker S and Dyer L (2003) Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48(1), 67–91. [https://doi.org/10.1016/S0749-596X\(02\)00515-6](https://doi.org/10.1016/S0749-596X(02)00515-6)
- Nieuwenhuis R, te Grotenhuis M and Pelzer B (2012) influence.ME: Tools for detecting influential data in mixed effects models. *The R Journal* 4(2), 38–47.
- Omaki A (2005) *Working memory and relative clause attachment in first and second language processing*. (Unpublished M.A. Thesis). University of Hawai'i.
- Omaki A (2016) Linking learning and parsing. *Bilingualism: Language and Cognition*, 1–2. <https://doi.org/10.1017/S1366728916000985>
- Pan HY, Schimke S and Felsler C (2015) Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism* 19(3), 298–313. <https://doi.org/10.1177/1367006913515769>
- Papadopoulou D and Clahsen H (2003) Parsing strategies in L1 and L2 sentence processing. *Studies in Second Language Acquisition* 25(4), 501–528. <https://doi.org/10.1017/S0272263103000214>
- Parker D, Shvartsman M and Van Dyke JA (2017) The cue-based retrieval theory of sentence comprehension: New findings and new challenges. In Escobar L, Torrens V and Parodi T (eds), *Language Processing and Disorders*. Newcastle: Cambridge Scholars Publishing, pp. 121–144.
- Pearlmutter NJ and MacDonald MC (1995) Individual Differences and Probabilistic Constraints in Syntactic Ambiguity Resolution. *Journal of Memory and Language*. <https://doi.org/10.1006/jmla.1995.1024>
- Pozzan L and Trueswell JC (2016) Second language processing and revision of garden-path sentences: A visual word study. *Bilingualism: Language and Cognition* 19(3), 636–643. <https://doi.org/10.1017/S1366728915000838>
- Ratcliff R (1993) Methods for dealing with reaction time outliers. *Psychological Bulletin* 114(3), 510–532.
- R Core Team. (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Rosenbach A (2008) Animacy and grammatical variation — Findings from English genitive variation. *Lingua* 118(2), 151–171. <https://doi.org/10.1016/j.lingua.2007.02.002>
- Sachs J (1967) Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics* 2, 437–442.
- Sachs J (1974) Memory in reading and listening to discourse. *Memory & Cognition* 2(1), 95–100.
- Sanchez CA, Wiley J, Miura TK, Colflesh GJH, Ricks TR, Jensen MS and Conway ARA (2010) Assessing working memory capacity in a non-native language. *Learning and Individual Differences* 20(5), 488–493. <https://doi.org/10.1016/j.lindif.2010.04.001>
- Sturt P (2003) The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language* 48(3), 542–562. [https://doi.org/10.1016/S0749-596X\(02\)00536-3](https://doi.org/10.1016/S0749-596X(02)00536-3)
- Swets B, Desmet T, Hambrick DZ and Ferreira F (2007) The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General* 136(1), 64–81. <https://doi.org/10.1037/0096-3445.136.1.64>
- Temple L (2000) Second language learner speech production. *Studia Linguistica* 54(2), 288–297. <https://doi.org/10.1111/1467-9582.00068>
- Traxler MJ (2002) Plausibility and subcategorization preference in children's processing of temporarily ambiguous sentences: Evidence from self-paced reading. *The Quarterly Journal of Experimental Psychology Section A*, 1(1), 75–96. <https://doi.org/10.1080/0272498014300017>
- Traxler MJ (2007) Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory and Cognition* 35(5), 1107–1121. <https://doi.org/10.3758/BF03193482>
- Traxler MJ, Williams RS, Blozis SA and Morris RK (2005) Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language* 53(2), 204–224. <https://doi.org/http://dx.doi.org/10.1016/j.jml.2005.02.010>
- Trueswell JC, Sekerina I, Hill NM and Logrip L (1999) The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition* 73, 89–134.
- Turner ML and Engle RW (1989) Is working memory capacity task dependent? *Journal of Memory and Language* 28, 127–154.
- Unsworth N, Heitz PR, Schrock CJ and Engle WR (2005) An automated version of the operation span task. *Behavior Research Methods* 37(3), 498.
- Van Dyke JA and McElree B (2011) Cue-dependent interference in comprehension. *Journal of Memory and Language* 65(3), 247–263. <https://doi.org/10.1016/j.jml.2011.05.002>
- Webman-Shafraan R and Fodor JD (2016) Phrase length and prosody in on-line ambiguity resolution. *Journal of Psycholinguistic Research* 45, 447–474. <https://doi.org/10.1007/s10936-015-9358-2>
- Witzel J, Witzel N and Nicol J (2012) Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics* 33(2), 419–456. <https://doi.org/10.1017/S0142716411000427>