Incrementality in Pre-nominal Relative Clause Attachment

Pavel Logačev1, Özgur Aydın2, & Müge Tuncer3

1. Boğaziçi University University, Istanbul, Turkey
2. Ankara University, Ankara, Turkey
3. Anadolu University, Eskişehir, Turkey

Abstract

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words “**here we show**” or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a more **general context**. Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

*Keywords:* keywords Word count: X

One important question in human sentence comprehension has always been how the human parser deals with ambiguity. A central phenomenon in that regard is the ambiguity advantage, first reported by Traxler, Pickering, and Clifton (1998), who found that sentences with globally ambiguous relative clause attachment such as (1a) were read faster than their locally ambiguous counterparts, such as (1b) and (1c). van Gompel, Pickering, and Traxler (2000) proposed a variable-choice two-stage model of ambiguity resolution to explain this finding: According to their unrestricted race model (URM), the parser uses all available

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line. Enter author note here.

Correspondence concerning this article should be addressed to Pavel Logačev, Postal address. E-mail:

pavel.logacev@boun.edu.tr

information to immediately resolve the attachment ambiguity as early as possible – in this case, at the first word of the relative clause in (1). It does so by attempting to compute all possible structures simultaneously when it encounters an ambiguity, and choosing the structure which is completed first. When faced with a relative clause (RC) attachment ambiguity, such as in (1), the parser tries to attach the relative clause to each of the two available noun phrases simultaneously: NP1 (headed by *son*) and NP2 (headed by \*driver). The URM further assumes that the time required to attach the relative clause to either noun phrase varies from trial to trial and depends on a variety of factors, such as the frequency of modification of NP1 and NP2, discourse-level information, and world knowledge. In consequence, when the parser encounters an ambiguity, it forms an N1 attachment when it is computed the fastest, or an N2 attachment when it is computed the fastest.

(1) a. globally ambiguous The son of the driver *that had the moustache* was pretty cool.

# b. locally ambiguous, N1 attachment c. locally ambiguous, N2 attachment

The car of the driver *that had the moustache* was pretty cool.

The driver of the car *that had the moustache* was pretty cool.

Because this local ambiguity is sometimes resolved towards one reading and sometimes towards the other in all attachment conditions, disambiguation of the local ambiguity at

*moustache* results in occasional reanalysis in both unambiguous conditions, (1b) and (1c). Since globally ambiguous sentences like (1a) are never disambiguated, reanalysis is never required. Therefore, on a certain proportion of occasions, locally ambiguous sentences cause reanalysis-related slowdowns, while globally ambiguous sentences do not, resulting in an ambiguity advantage. The ambiguity advantage effect has been replicated in multiple experiments Dillon, Andrews, Rotello, & Wagers (2019).

Multiple alternative explanations of the ambiguity advantage, such as Surprisal (Levy, 2008), and constraint-based models (Green & Mitchell, 2006; Vosse & Kempen, 2009) have been proposed. Importantly, there are alternative explanations for this finding. For instance, Levy’s (2008) independently motivated surprisal account (cf. also Hale (2001)) posits that the processing difficulty at each word is proportional to the word’s surprisal value, that is, to its negative conditional log-probability, given the context. According to the surprisal account, the potentially disambiguating word “moustache” is associated with higher surprisal in locally ambiguous sentences such as (1b) and (1c) than in globally ambiguous sentences such as (1a) because it is compatible with both possible structures in ambiguous sentences, but with only one structure in unambiguous sentences. Thus, the potentially disambiguating word has a higher conditional probability in ambiguous sentences than in unambiguous sentences. Because higher conditional probability corresponds to lower surprisal, the lower reading times in globally ambiguous conditions compared to the locally ambiguous conditions are explained by the lower surprisal in ambiguous sentences.

Another influential alternative account has been the underspecification account

(Swets, Desmet, Clifton, & Ferreira, 2008), which posits that the parser may underspecify ambiguities if their disambiguation is not required by the task. They argue that because participants in the studies by Traxler et al., as well as Van Gompel et al. were asked fairly superficial questions about the sentences they read, they decided that ambiguous sentences don’t need to be disambiguated as they knew that they wouldn’t be asked comprehension questions about the ambiguous grammatical relation. In other words, readers underspecify RC attachment when they believe that it is irrelevant. For example, RCs with ambiguous attachment such as in sentence (1a) are not attached to NP1 or NP2, but instead are associated with the entire complex NP indicating attachment to any constituent within the complex NP. Swets et al. further postulate that underspecification of ambiguous relative clause attachment saves time, which explains the ambiguity advantage attested in the abovementioned studies. Importantly, underspecification is assumed to be a viable strategy only when tasks demands do not require disambiguation. Swets et al. tested this explanation by conducting a self-paced reading study in which participants read sentences such as in (2) and were asked either (i) superficial questions such as *‘Was anyone humiliated/proud?’* or (ii) questions about the RC attachment ambiguity *‘Did the maid/princess/son scratch in public?’*. They found an ambiguity advantage effect in the reading time data of participants in the superficial questions group, while the RC questions group did not display such an effect. Based on these findings, Swets et al. argue that the type of questions asked modulates task demands and thus switches underspecification on and off.

# (2) a. globally ambiguous b. locally ambiguous, N1 attachment c. locally ambiguous, N2 attachment

The maid of the princess who scratched herself in public was terribly humiliated.

The son of the princess who scratched himself in public was terribly humiliated.

The son of the princess who scratched herself in public was terribly humiliated.

While Swets et al.’s findings do constitute evidence for a certain degree of influence of task-demands in reading, they do not necessarily constitute evidence for underspecification. Logačev and Vasishth (2016) have demonstrated that Swets et al.’s findings are compatible with a race-based account as well, under the assumption that relative clauses are attached only after they have been processed, which is also a necessary assumption of the underspecification account. According to their proposal, called *SMCM* (*stochastic multiple channel model*), the ambiguity advantage arises as a result of so-called statistical facilitation (Raab (1962)) due to a race between two independent attachment processes. Because ambiguous sentences have two possible readings while unambiguous sentences have only one, a race between two independent attachment processes executed in parallel leads to a decrease in average processing times because the probability that a race between two independent processes of approximately the same speed terminates quickly is higher than the probability that a specific process will do so. Using a computational implementation of the SMCM model, Logačev and Vasishth demonstrate that the Swets et al. findings are compatible with a race model like the SMCM, under the assumption that task demands have differential effects on preferred and dispreferred relative clause attachment times. The latter assumption is supported by Swets et al.’s finding that the difference in reading times between sentences with preferred and dispreferred unambiguous attachment was bigger in the RC questions condition than in the superficial questions condition. While the mechanism causing this interaction between task demands and attachment remains unclear, Swets et al.’s findings appear to be compatible with models that do not assume a modulation of the ambiguity resolution strategy by task demands. Importantly, both accounts of Swets et al.’s findings (the underspecification account and the SMCM) must assume some mechanism by which task demands affect either the duration or the quality of some parsing operations, even in unambiguous sentences. While we are not aware of any independent evidence for an interaction between sentence complexity and task demands, it is in line with prior research demonstrating the influence of task demands on reading Dempsey & Brehm (2020).

According to the underspecification account, the parser chooses to underspecify ambiguous sentences whenever doing so does not conflict with carrying out what is perceived as the main experimental task, such as answering comprehension questions. Importantly, the underspecification mechanism could function in one of two ways: One possibility is that, task demands permitting, the parser routinely underspecifies RC attachment ambiguities when it first encounters them, i.e., at the relativizer. Later, it disambiguates them, when that is deemed necessary due to task demands, or if the ambiguity is disambiguated. We will refer to this mechanism as *early underspecification*. A second possibility is that the parser delays RC attachment entirely until it has processed the key parts of the RC (at a minimum, the verb and its core arguments). Having done so, it decides whether to attach or to underspecify based on whether the RC is ambiguous and on the task demands. We will refer to this mechanism as *late underspecification*. The key difference between the two accounts is that according to early underspecification an underspecified representation is always created as soon as possible and may be disambiguated at a later point, while according to late underspecification, structure-building is delayed until the parser decides to either underspecify or to disambiguate.

While both underspecification accounts and the SMCM make the same predictions for the processing of post-nominal relative clauses, such as in English in (1) and (2), their predictions diverge for post-nominal relative clauses, such as in Turkish.

# Ambiguity resolution in pre- and post-nominal relative clauses

In the sentences in (1) and (2), both, race models as well as the underspecification account predict the occurrence of the ambiguity advantage at the relative clause. This is because relative clauses in English are post-nominal, i.e., they follow the noun phrases they modify. As a result, when readers have processed a post-nominal relative clause to a sufficient degree to attach it, both potential attachment sites are immediately available, allowing the parser to either start a race, or to underspecify. The situation is quite different in languages with pre-nominal relative clauses, such as Turkish (Göksel and Kerslake (2005)). In constructions like (3), the potential attachment sites become available *successively* after the relative clause has been processed. For example, in the globally ambiguous sentence in (3), the relative clause *‘who cried’* modifies either the complex subject noun phrase headed by *aunt* (N2), or its genitive possessor *girl* (N1). Thus, this sentence can either mean that the girl had cried (local attachment; to the NP headed by N1), or that the girl’s aunt had cried (non-local attachment; to the NP headed by N2). For the sake of clarity, we will refer to these attachment conditions as N1 and N2 attachment, i.e., in reference to the position of the NP’s head noun.[[1]](#footnote-2)

(3) [Sokakta ağlayan]*RC* kızın halası evde arkadaşlarını bekledi. on the street cried the girl’s aunt at home her friends waited.

‘The aunt of the girl who had cried on the street waited for her friends at home.’

The earliest point in this sentence at which an attachment can be made is the first noun (*girl*). This is because after reading this noun, the parser has processed a relative clause, as well as a potential attachment site for it. Therefore, a dependency between *girl* and *‘who had cried’* can be established. However, the parser does not yet know whether the sentence is ambiguous, but importantly, it has good reason to believe that it might be. This is because the genitive case suffix on the first noun phrase signals that it is likely the possessor of a complex noun phrase and that the parser can thus expect another potential attachment site at the next word.

— Prior research on Turkish on attachment preference and possibly relative difficulty of N1/N2 attachment (Kirkici, Nazik, Ankara1, Ankara2) —

The fact that the two possible attachment sites are encountered successively makes this structure particularly interesting for testing the idea that the parser strategically underspecifies relative clause attachment in order to save time. Under superficial task demands, according to early underspecification, the parser should always underspecify RC attachment in ambiguous and N1 sentences upon encountering N1 (but not in N2 conditions), and

XXX

This is because the optimal time-saving strategy in processing this structure would be underspecification of attachment when N1 is encountered. The next step in ambiguous sentences, when N1 is followed by an N2 to which the RC can also attach, the parser can maintain the underspecified structure.

When readings become available successively,

the SMCM predicts that attachment to N1 will be initiated at N1, and that an N1 reading should be computed on most, if not all, trials – unless attachment is postponed till N2

early underspecification predicts underspecification at N1 when possible, followed by disambiguation at N2 in the N1 attachment condition, which means there should be an ambiguity advantage at N2 in the due to the absence of attachment in the ambiguous condition.

* late underspecification if the parser waits just for the minimal set of arguments: identical to SMCM – N1 attachment
* late underspecification if the parser waits for the end of the possible attachment domain: underspecification due to ambiguity

In unambiguous sentences, when N1 is followed by an N2 which is incompatible with the

RC (or other material to which the RC cannot attach), the parser can fully specify RC attachment at this point. As a result, the strategic underspecification account predicts an ambiguity advantage in the processing of structures with pre-nominal relative clauses like

(3).

The SMCM, on the other hand, assumes that the parser always computes at least one attachment as soon as it can. There is no mechanism by which RC attachment can be delayed. As a result, it predicts that when task demands do not require the computation of both interpretations, the processes that will unfold in ambiguous sentences is the same as in the N1 attachment sentences, and thus does not predict an ambiguity advantage.

We conducted two experiments in order to test these predictions. In both experiments, task demands were superficial. In experiment 1, we disambiguated RC attachment by means of animacy, while in experiment 2 RC attachment was disambiguated morphosyntactically.

# Experiment 1

For this experiment, we created three types of sentences exemplified in (4a,b,c). Which NPs were available as attachment attachment sites was manipulated by means of animacy of N1 and N2: For example, in sentence (4b) the relative clause “who was shouting in the school building” can attach only to the noun *principal*, but not to *voice*, whereas in sentence (4c) the relative clause *‘who was fired’* can attach only to the noun *cook* but not to *yacht*. In (4a), on the other hand, both nouns are potential agents of crying, and therefore RC attachment is ambiguous.

## (4) a. ambiguous attachment

[Sokakta ağlayan]*RC* kızın halası evde arkadaşlarını bekledi. on the street cried the girl’s aunt at home her friends waited.

‘The aunt of the girl who had cried on the street waited for her friends at home.’

## b. unambiguous, N1 attachment

[Okulda bağıran]*RC* müdürün sesi caddenin karşısında

in the school building shouting principal voice street.gen opposite duyuldu. hear.

‘The voice of the principal who was shouting in the school building was audible from across the street.’

## c. unambiguous, N2 attachment

[İşten çıkarılan]*RC* yatın aşçısı işverenden parasını from work dismissed yacht.gen cook from employer his money almadı. did not take

‘The yacht’s cook who was fired by his employer did not take his money.’

According to the SMCM, the parser should attempt RC attachment while at N1, which would successfully terminate in the N1 attachment and in the ambiguous attachment conditions. At N2, no additional RC attachment should be carried out because the occasional comprehension questions in this experiment should not encourage the computation of a second interpretation of the sentence. In the N2 attachment condition, in contrast, the attempt at RC attachment at N1 will fail. At N2, RC attachment would be carried out successfully, resulting in a slowdown relative to the other two questions. It’s not clear whether the parser’s failure to attach the RC to N2 should lead to a speedup, a slowdown, or no appreciable difference in reading times.

The underspecification account makes one of two predictions. Under the assumption that the parser strategically uses underspecification to avoid unnecessary processing, we would expect it to always underspecify RC attachment during the processing of N1, in case it turns out that it doesn’t attach at all. Then, at N2, RC attachment should be carried out in NP1 and N2 attachment conditions. In the ambiguous condition, no such thing is expected to happen since there is no need to disambiguate the sentence when comprehension questions are occasional, as in this experiment. As a result, we expect no difference in reading times at N1, since we underspecify in all conditions, and an ambiguity advantage at N2, since the parser will attach the RC at N2 in both unambiguous conditions but not in the ambiguous condition. Moreover, we may expect a difference between the two unambiguous conditions.

Under the assumption that the parser can’t routinely underspecify RC attachment in anticipation of an ambiguity, and can only do so if both attachment sites are available, the underspecification account’s prediction match the predictions of the SMCM: early attachment at N1 in the N1 attachment and ambiguous conditions, and late attachment at N2 in the N2 attachment condition.

The underspecification account predicts a speed-up in the ambiguous condition at the second noun, compared to the local attachment condition. This prediction translates into an interaction between modifier type and the grammatical number of the second noun. Reading should slow down in RC attachment conditions when the second noun is singular. The URM makes predictions concerning the same interaction. Its prediction about the lack of an ambiguity advantage in Turkish translates into a predicted lack of an interaction. In other words, there should not be any significant interaction between modifier type and the grammatical number of the second noun in the RC attachment conditions.

Importantly, because the URM predicts no significant attachment-related speed-up; in other words, a statistical null-result is expected under this account. In order to differentiate between possible outcomes compatible with the URM and strategic underspecification, respectively, we need to quantify the expected magnitude of an ambiguity advantage. Swets, Desmet, Clifton, and Ferreira (2008) is to the best of our knowledge the only study which has demonstrated an ambiguity advantage in self-paced reading. They found ambiguity advantage effects of 52*ms* and 36*ms* in the superficial and occasional questions conditions respectively, though the latter was not significant. Three other studies have found evidence for this effect in eye tracking experiments: the original Traxler, Pickering, and Clifton (1998) article presented ambiguity-related speedups between 92*ms* and 32*ms* in total reading times at the disambiguating region in two eye tracking experiments. Both experiments involved relative clause attachment. In further two experiments concerning attachment of prepositional phrases, van Gompel, Pickering, and Traxler (2001) presented effects between 41*ms* and 122*ms*, mostly in total reading times, and one in regression-path duration. Furthermore, van **Table 1**

*Average word length by condition and position in the sentence. We considered the positions pre-critical (’shouting’), noun 1 (’principal’), noun 2 (’voice’), and spill-over (’street’).*

*(Standard deviations and range in parentheses.)*

|  |  |  |  |
| --- | --- | --- | --- |
| pre-critical | noun 1 | noun 2 | spillover |
| 7.4 (1.2; 6-10) | 5.8 (1.1; 4-7) | 5.4 (1.0; 4-7) | 5.5 (2.0; 3-9) |
| 6.7 (0.9; 5-8) | 5.7 (1.2; 4-8) | 5.3 (1.2; 3-7) | 5.5 (1.7; 3-8) |
| 7.8 (1.1; 7-10) | 5.7 (1.2; 3-7) | 5.3 (1.2; 3-7) | 5.7 (3.1; 3-13) |

Gompel, Pickering, Pearson, and Liversedge (2005) demonstrated speedups between 26ms and 86ms in total reading time and regression-path durations. This pattern suggests that most ambiguity advantage effects appear to concentrate somewhere around 50*ms*. Thus, under the strategic underspecification account, we expect an attachment-related speedup of approximately 50*ms*. Under the SMCM, we expect no such speed-up.

The key question we wanted to answer with this experiment is whether there is an ambiguity advantage in Turkish, under task demands that have produced one in English. We operationalized an ambiguity advantage as the observation of a speed-up in the ambiguous condition relative to *both* unambiguous conditions and its magnitude as the difference between the fastest unambiguous condition and the ambiguous condition. Since the smallest previously observed ambiguity advantage effect in prior literature (Swets, Desmet, Clifton, & Ferreira, 2008; Traxler, Pickering, & Clifton, 1998; van Gompel, Pickering, Pearson, & Liversedge, 2005; van Gompel, Pickering, & Traxler, 2001) amounted to a 26*ms* difference between the preferred unambiguous condition and the ambiguous condition (range: 26*ms −* 122*ms*), we expected an effect of at least that magnitude.

# Method

## Materials

We constructed 36 experimental sentences such as in (4), i.e., twelve for each attachment condition. Disambiguation was effected by means of animacy. Relative clauses always comprised two words. The critical region and the spillover region were approximately equal in length across the three conditions: (cf. Table 1).

## Participants

A group of 39 native Turkish speakers aged 18–28 (mean = 23; SD = 3.2), who were students at Ankara University (in Ankara, Turkey) at the time of testing, participated in the study. All of the participants reported to have normal or corrected to normal vision; none of them reported to have extensive stays abroad or neurological/psychiatric disorder that may impact language comprehension ability in Turkish. The participants were informed that their participation was voluntary and were asked to give their consent allowing us to anonymously process their experimental data for scientific purposes. Participants provided informed consent and the procedures in this study were compliant with the Ankara University research ethics requirements as well as with the ethical principles outlined in the Helsinki Declaration on research involving human subjects.

## Procedure

All the experiments took place at Ankara University Linguistics Laboratory [(http: //dilab.ankara.edu.tr/)](http://dilab.ankara.edu.tr/), with each participant individually in a single session. The participants were required to perform the sentence processing task in a dimly lit room and seated within 60 cm distance from a 1680 x 1050 pixels 22-inch LCD desktop monitor with 60 Hz refresh rate. During the performance of the sentence processing task, the eye-tracker device (RED 500 by Sensor Motoric Instruments, SMI) with chin rest recorded the participants’ binocular eyes movements at a sampling rate of 500*Hz*, but only the left eye was analyzed. Before beginning the sentence processing task, participants were told that they would read Turkish sentences and answer comprehension questions on the computer screen while the eye-tracker recorded their eye movements. Each participant was given a practice session to get used to the dynamics of the task. A 5-point calibration was executed before and after the practice. During the testing phase, each participant read forty-eight sentences (36 test sentences and 12 filler sentences) in two sentence blocks of twenty-four sentences each. Before each sentence, a fixation cross appeared for 500 ms trigger duration, in order to help the participants to fixate their eyes in the initial point of each sentence on the screen. The sentences appeared on the screen one at a time in one line in a grey background. The degree of angle per character was 0.074 cm (Courier New, 36 point). After every two sentences, superficial questions or questions questions about the RC attachment ambiguity related to the last sentence read was presented, the answer to which was given by clicking on one of two responses shown on the screen. Each session lasted about 25 minutes.

## Statistical Analysis

We excluded all fixations which were shorter than 80 ms or longer than 1200 ms prior to using the *em2* package (Logačev & Vasishth, 2013) in R [Version 4.0.4; R Core Team (2018)] for computing reading time measures from the eye movement record. We used the *tidyverse* and *ggplot2* packages (Wickham, Averick, Bryan, Chang, McGowan, François, Grolemund, Hayes, Henry, Hester, Kuhn, Pedersen, Miller, Bache, Müller, Ooms, Robinson, Seidel, Spinu, Takahashi, Vaughan, Wilke, Woo, and Yutani, 2019; Wickham, 2016) for data processing and plotting, and the R packages *brms* (Bürkner, 2017; Bürkner, 2018) and *rstan* (Stan Development Team, 2020) to fit Bayesian generalized hierarchical linear models (e.g., Gelman & Hill, 2007; Kruschke, 2015; McElreath, 2016; Vasishth, Nicenboim, Beckman, Li, & Kong, 2019) to the eye-tracking measures of interest. The discrete dependent measures regression probability and probability of refixation were modeled with generalized hierarchical linear models with a logit-link. All reading time measures were modeled using generalized hierarchical linear models assuming log-normally distributed residuals. We used normal priors for all coefficients for both, log-normal models (*µ* = 0, *σ* = 0*.*2), as well as logit models (*µ* = 0, *σ* = 1). All models were fitted using four chains with 2*,*500 post-warmup samples each.

The models presented below included fixed effects of attachment coded using orthogonal simple contrasts (comparing each of the unambiguous attachment condition to the ambiguous condition), as well as centered log word length as predictors. Word length was used as a covariate in order to account for any differences in critical word length between our experimental sentences. We included random intercepts for participants and items, as well as maximal by-participant random slopes and correlations between random effects.

While all coefficients for reading time measures were estimated using sum contrasts on a log-scale, we used the model’s posterior MCMC samples to construct 95% credible intervals for the pairwise differences between each of the unambiguous conditions and the ambiguous condition on the original scale (in *milliseconds*). Doing so allowed us to compare the size of the effects we found with previously attested instances of the ambiguity advantage.

# Results

The average comprehension questions accuracy was 78*.*8. Figures 1 show a summary of the eye-tracking measures computed for the the critical regions: (i) *First-pass reading time* (FPRT), which is the amount of time spent reading a word for the first time when that the first fixation on that word was *progressive* (i.e., the reader has not yet read anything to the right of this word). (ii) *Regression-path duration* (RPD), which is the amount of time spent on the word starting with a progressive first fixation, until the gaze moves to the right of that word. (iii) *Total fixation time* (TFT), which is the total amount of time spent fixating the critical word, including re-reading. (iv) The percentage of *first-pass regressions*, which is the proportion of trials on which readers regress after reading the critical word).

Figure 2 shows the estimated differences between the unambiguous conditions and the ambiguous condition for each word position and eye-tracking measure. It also shows the upper and lower boundaries of the 90% credible interval for each position (noun 1, noun 2, spill-over region).

## N2 attachment

The estimates show a moderate amount of evidence for a slowdown in N2 attachment conditions in first-pass reading times (*N2:* CrI [-7; 49], P(*β* > 0) = .89; *spill-over:* CrI [-11; 44], P(*β* > 0) = .84), with stronger evidence for a slowdown in regression-path duration (*N2:* CrI [18; 109], P(*β* > 0) = .99, *spill-over:* CrI [-4; 73], P(*β* > 0) = .93). In total fixation time, we found a slowdown in N2 attachment conditions at N2 (CrI [-3; 105], P(*β* > 0) = .94), and a weaker effect in the same direction at the (CrI [-22; 68], P(*β* > 0) = .81). We further found an N2 attachment-related slowdown in regression probability at N2 (CrI [0.6;

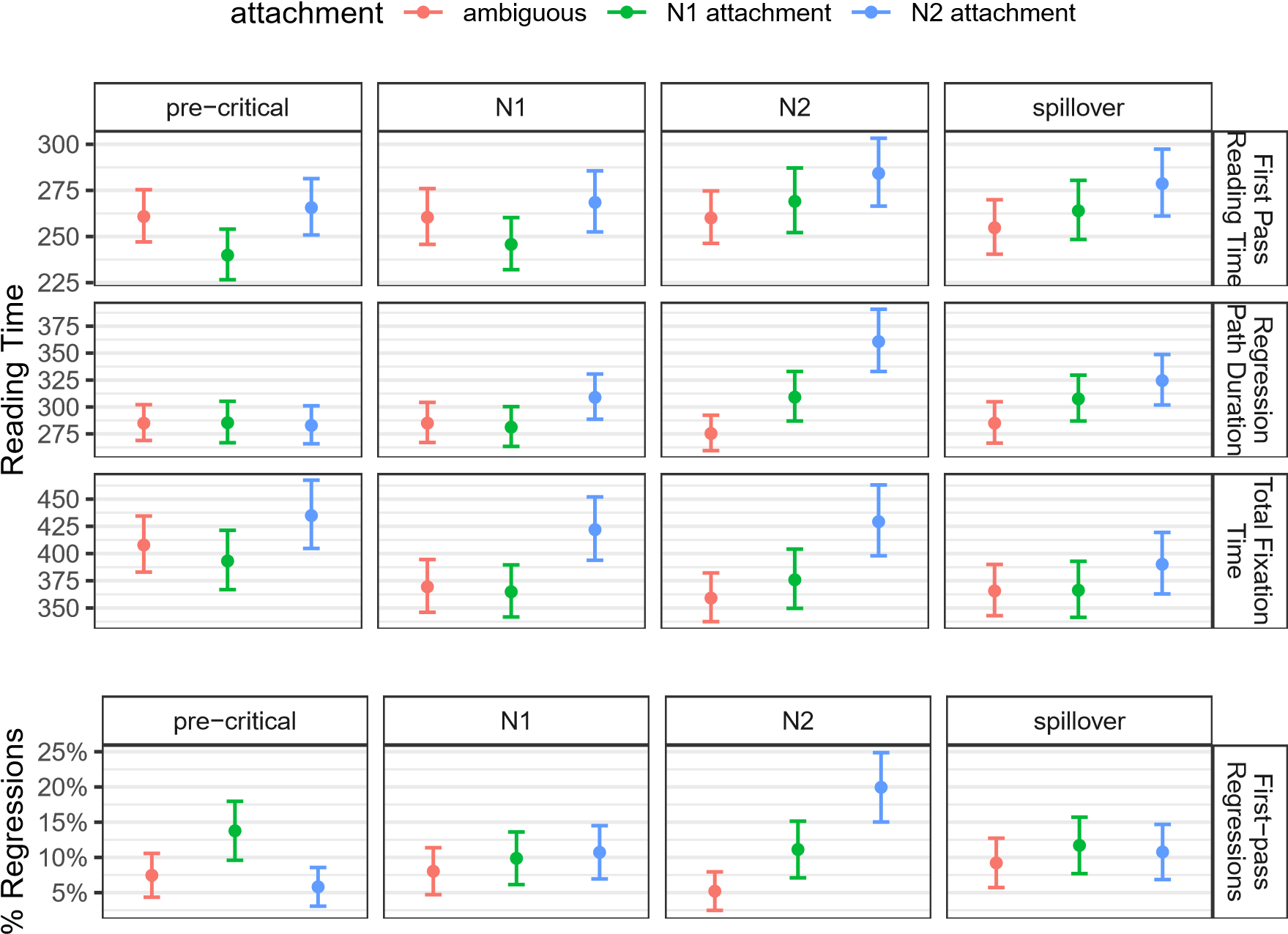
2.0], P(*β* > 0) = .998).

## N1 attachment

While most eye tracking measures on most regions reflected an N2 attachment-related slowdown, only regression path duration and regression probability suggested a slowdown in N1 attachment conditions: RPD at N2 (CrI [-25.7; 56.3], P(*β* > 0) = .72) and at the spill-over region (CrI [-20.9; 52.1], P(*β* > 0) = .75), as well as regression probability at N2 (CrI [-0.3; 1.2], P(*β* > 0) = .86).

## Ambiguity Advantage

The posterior probability of the difference between N1 attachment and the ambiguous being larger than the smallest observed ambiguity advantage so far was *P*(*β ≥* 26*ms*) = 0*.*32,



# Figure 1

*Average first-pass regression probabilities and backtransformed log-reading time measures FPRT, RPD, and TFT for the critical regions by condition. Within-participants standarderrors in brackets (Cousineau, 2005; Morey, 2008).*

while the posterior probability of the difference between N2 attachment and the ambiguous being larger than the smallest observed ambiguity advantage so far was *P*(*β ≥* 26*ms*) = 0*.*91.

# Discussion

We found substantial processing difficulty in the N2 attachment condition relative to the ambiguous attachment condition, while the effect of unambiguous N1 attachment was less clear. In the N2 attachment condition, we found indications of processing difficulty in most eyetracking measures starting with regression path duration on N1, suggesting that reading in N2 attachment conditions is initially slowed by the the fact that the RC cannot be attached to the first structurally available noun phrase, headed by N1. Subsequently, it may be slowed down relative to the ambiguous baseline due to RC attachment to the noun phrase headed by N2 which unfolds at N1, as this process would arguable not be carried out in the N1 attachment and ambiguous conditions.

The relative difficulty of the N1 attachment sentences is less clear. While there was

First−pass reading time

N1 attachment − ambiguous

noun 1

noun 2

spill−over

N2 attachment − ambiguous

−25 0 25 50 −25 0 25 50 −25 0 25 50

Estimate (ms)

Regression−path duration

N1 attachment − ambiguous

noun 1

noun 2

spill−over

N2 attachment − ambiguous

−40 0 40 80 −40 0 40 80 −40 0 40 80

Estimate (ms)

Total fixation time

N1 attachment − ambiguous

noun 1

noun 2

spill−over

N2 attachment − ambiguous

−40 0 40 80 120 −40 0 40 80 120 −40 0 40 80 120

Estimate (ms)

% Regressions

N1 attachment − ambiguous

noun 1

noun 2

spill−over

N2 attachment − ambiguous

0 1 2 0 1 2 0 1 2

Estimate (log−odds)

# Figure 2

*Estimates and 90% credible intervals (thin lines) as well as 80% credible intervals (thick lines) for the analyses of first-pass reading time, regression-path duration, total fixation time, regression probability, and refixation probability. The posterior probability that the parameter is smaller than (or larger than) zero is larger than 90% if the 80% credible interval excludes zero, and larger than 95% if the 90% credible interval excludes zero.*

no appreciable slowdown in first-pass reading time or total fixation time at N2 in the N1 attachment condition relative to the ambiguous condition, the regression-path duration as well the regression probability indicated somewhat higher processing difficulty in the N1 attachment conditions compared to the ambiguous conditions. Since the posterior probability of a slowdown is 0*.*72, while the posterior probability of the slowdown being at least as large as the smallest previously observed ambiguity advantage was 0*.*32, the results of this experiment do not provide unambiguous evidence for or against the presence of an ambiguity advantage in Turkish. However, they do suggest that it is more likely that there is no ambiguity advantage of a previously attested magnitude than that there is one.

# Experiment 2

Our second experiment was designed to address two issues which may complicate the interpretation of the first experiment: Firstly, the analysis of multiple eye tracking measures at multiple regions increases the chances that one of them exhibits a theoretically interesting effect due to chance, even if there is no underlying effect (e.g, (**BernhardTitus:2015?**)). Secondly, it is possible we hypothesized that a plausibility-based disambiguation based on manipulation maybe plausibility isn’t a strong enough manipulation. ... this design made a within-items(?) design?

The aim of the second experiment was to confirm the findings of the first experiment using a different design. The critical regions in the experimental sentences in this experiment had a structure similar to sentence (5), in which the head noun of the relative clause (*football player*/*fan*) performs the function of the subject of the embedded verb (*hit*). The grammatical object of all relative clauses was the reciprocal pronoun *each other*. RC attachment was disambiguated by the grammatical number of the nouns. In (5), for instance, the relative clause *‘hit each other’* can attach to the NP headed by *fans* because it is plural, but not to the NP headed by *football player*, because it is singular and therefore cannot bind the reciprocal *each other*.

(5) [Birbirini döven]*RC* futbolcu-nun hayran-lar-ı terk etti.

each other hit footballer.sg-gen fan-pl-poss leave did.

*‘The fan(s) of the football player(s) who hit each other left the stadium.’*

We used the fact that relative clauses with a reciprocal as a grammatical object can only attach to plural noun phrases to create three attachment conditions: N1 attachment, N2 attachment and ambiguous attachment, as illustrated in (6). Because plurals were formed through addition of the plural suffix *-ler/-lar*, reading times for plural and singular versions of a word are not easily comparable. Thus, in order to control for possible effects of length, we created three control conditions illustrated in (7) in which the complex noun phrase was modified by an adjective which could only attach to N1. In the examples in (7) it is *‘Fenerbahçeli’* (meaning *‘of Fenerbahçe’*, a well-known Turkish football team). These additional conditions will serve as a baseline, any differences from which must be considered effects of RC attachment.

In this self-paced reading experiment, we asked superficial comprehension questions occasionally. Questions were kept superficial in order to parallel the task demands employed by previous experiments which demonstrated an ambiguity advantage (Swets, Desmet, Clifton, & Ferreira, 2008; Traxler, Pickering, & Clifton, 1998; van Gompel, Pickering, Pearson, & Liversedge, 2005; van Gompel, Pickering, & Traxler, 2001).

(6) Dün akşam, ...

Yesterday evening,

## a. ambiguous attachment (plural-plural)

[birbirini döven]*RC* futbolcu-lar-ın hayran-lar-ı each other hit football player-pl-gen fan-pl-poss

## b. N1 attachment (plural-singular)

[birbirini döven]*RC* futbolcu-lar-ın hayran-ı each other hit football player-pl-gen fan.sg-poss

## c. N2 attachment (singular-plural)

[birbirini döven]*RC* futbolcu-nun hayran-lar-ı each other hit football player.sg-gen fan-pl-poss

... stadyumu hemen terk etti. stadium immediately leave did.

*‘The fan(s) of the football player(s) who hit each other left the stadium immediately, yesterday evening.’*

(7) Dün akşam, ...

Yesterday evening,

## a. ambiguous attachment control (plural-plural)

[Fenerbahçe-li]*ADJ* futbolcu-lar-ın hayran-lar-ı

Fenerbahce-adj hit football player-pl-gen fan-pl-poss

## b. N1 attachment control (plural-singular)

[Fenerbahçe-li]*ADJ* futbolcu-lar-ın hayran-ı

Fenerbahce-adj hit football player-pl-gen fan.sg-poss

## c. N2 attachment control (singular-plural)

[Fenerbahçe-li]*ADJ* futbolcu-nun hayran-lar-ı

Fenerbahce-adj hit football player.sg-gen fan-pl-poss

... stadyumu hemen terk etti. stadium immediately leave did.

*‘The fan(s) of the Fenerbahce football player(s) left the stadium immediately, yesterday evening.’*

# Method

## Materials

Forty-two experimental sentence sets like (6) and (7) were constructed. All relative clauses comprised between two and three words and always started with a reciprocal pronoun. Sentences were divided into different lists according to a latin-square design, such that every participant read exactly one sentence from each sentence set, and seven sentences from every condition. Experimental sentences were mixed with 65 filler sentences, so that every participant read 107 sentences during the course of the experiment. Each list was randomized prior to presentation.

## Participants

Thirty-six students of Anadolu University in Eskişehir, Turkey participated in the experiment. All participants were native speakers of Turkish; their age range was 19-29 years. One experimental session took approximately 40 minutes to complete. Participants were paid 15 Turkish Lira for participation. Participants provided informed consent and the procedures in this study were compliant with the Anadolu University research ethics requirements as well as with the ethical principles outlined in the Helsinki Declaration on research involving human subjects.

## Procedure

The task was self-paced non-cumulative word-by-word reading. Presentation and recording was done with the Linger software package, version 2.94 by Doug Rohde. At the beginning of a trial all words, masked by underscores, appeared. Participants pressed the space bar to reveal the next word. As the next word appeared, the current one was masked by underscores again. The time between key-presses was recorded as the reading time for the word. Each participant read five practice sentences before the start of the experiment. Questions about the sentence were asked on one third of all trials (14 questions about experimental sentences, 21 questions about fillers). Participants answered questions with yes' orno’ by pressing the corresponding button on the keyboard.

## Statistical Analysis

We used the *tidyverse* and *ggplot2* packages (Wickham, Averick, Bryan, et al., 2019;

Wickham, 2016) for data processing and plotting, and the R packages *brms* (Bürkner, 2017; Bürkner, 2018) and *rstan* (Stan Development Team, 2020). All reading time measures were modeled using Bayesian generalized hierarchical linear models assuming log-normally distributed residuals (e.g., Gelman & Hill, 2007; Kruschke, 2015; McElreath, 2016; Vasishth, Nicenboim, Beckman, Li, & Kong, 2019). We used normal priors for all coefficients (*µ* = 0, *σ* = 0*.*2). All models were fitted using four chains with 2500 post-warmup samples each.

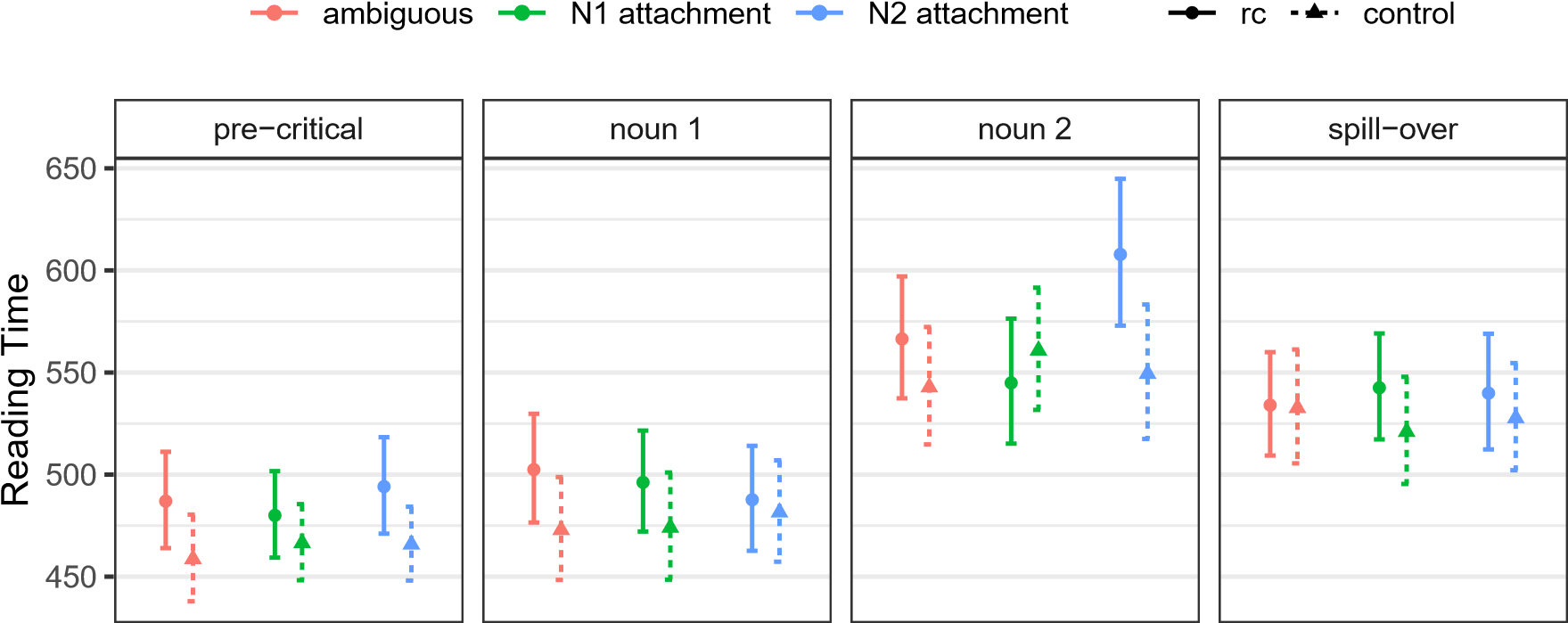
All models included fixed effects of (i) centered log-word length, (ii) the availability of

N1 and N2 as head nouns for RC attachment, (iii) the presence of a relative clause (*relative clause* vs. *control*), as well as (iv) the interactions between the presence of a relative clause and the availability of N1 and N2 as head nouns for RC attachment. Word length was used as a covariate in order to account for any differences in critical word length between our experimental sentences. For word length at critical regions N1 and N2, the word length of the singular form was used. All contrasts were centered. We included varying intercepts for participants and items, as well as maximal by-participant and near-maximal by-item slopes, as well as correlations between random intercepts and slopes. By-item random slopes of word length were not included in the models for critical regions N1 and N2, because they would be colinear with the effect of the grammatical number of those nouns.

The availability of N1 and N2 as head nouns for RC attachment was coded using (orthogonal) simple contrasts comparing (i) the average of (6b) and (7b) to the average of (6a) and (7a), as well as (ii) the average of (6c) and (7c) to the average of (6a) and (7a). Because the coefficients associated with the interaction contrasts represent the differences in the effect of N1 and N2 availability associated with the presence of a relative clause, they can be interpreted as the respective effects of unambiguous N1 and N2 attachment, relative to the ambiguous condition.

While all coefficients for reading time measures were on a log-scale, we used the model’s posterior MCMC samples to construct 95% credible intervals on the original scale (in *milliseconds*) for (i) the main effect of relative clause attachment, as well as for (ii) the interactions between N1 and N2 unvailability and the presence of a relative clause, which correspond to the pairwise differences between each of the unambiguous conditions and the ambiguous condition. Doing so allowed us to compare the size of the effects we found with previously attested instances of the ambiguity advantage.

# Results



# Figure 3

*Average backtransformed log-reading times for the critical regions by condition. Withinparticipants 95%-confidence intervals in brackets. (Cousineau, 2005; Morey, 2008)*

Figure 3 shows the average log-reading times transformed back to the original scale in milliseconds and figure 4 shows the estimates and credible intervals for the key contrasts at the critical regions in milliseconds. The effects of attachment shown in it correspond to the interactions between the presence of a relative clause and the contrasts encoding the grammatical number of N1 and N2, thus accounting for the effect of RC attachment when a relative clause is present. Figure 5 in the appendix shows all model coefficients in the original units (on the log-scale).

As figure 4 shows, we found moderate evidence for a slowdown in relative clause attachment conditions relative to the control conditions which occurred at both, N1 (CrI [-2;

noun 1

noun 2

spill−over

N2 attachment − ambiguous

N1 attachment − ambiguous

RC attachment − control

−50 0 50 100 −50 0 50 100 −50 0 50 100

Estimate (ms)

# Figure 4

*Estimates and 90% credible intervals (thin lines) as well as 80% credible intervals (thick lines) for the effect (in milliseconds) of unambiguous RC attachment relative to ambiguous attachment on reading times at the critical regions. The posterior probability that the parameter is smaller than (or larger than) zero is larger than 90% if the 80% credible interval*

*excludes zero, and larger than 95% if the 90% credible interval excludes zero.*

39], P(*β* > 0) = .93) and N2 (CrI [-4; 49], P(*β* > 0) = .92). In the *N1 attachment condition*, relative to the ambiguous condition, we found moderate evidence for a speedup at N2 (CrI [-85; 9], P(*β* < 0) = .91), which did not persist until the the spill-over region (CrI [-26; 65], P(*β* < 0) = .25). Meanwhile, there was moderate evidence of a slowdown in the *N2 attachment condition* relative to the ambiguous condition at N2 (CrI [-28; 94], P(*β* > 0) = .81) and nearly no indication of such an effect at the spill-over region (CrI [-35; 57], P(*β* > 0) = .65).

The posterior probability that the ambiguity advantage effect (the difference between the N1 attachment and the ambiguous condition) was larger than the smallest ambiguity effect attested in the literature was *P*(*β ≥* 26*ms*) = 0*.*01 at N2, and *P*(*β ≥* 26*ms*) = 0*.*39 at the spill-over region.

# Discussion

We found an across-the-board slowdown in relative clause attachment conditions compared to the control conditions, which likely reflects the increased processing difficulty associated with the processing and attachment of relative clauses, compared to adjectives. While we did not find strong evidence for a slowdown at N2 in N2 attachment sentences, the finding of a slowdown in the range of [-41; 107] (90% CrI) was consistent with the slowdown in N2 attachment sentences in RPD at N2 in experiment 1, which was in the range of [-41; 107] (90% CrI). As in experiment 1, this slowdown likely reflects late RC attachment in the N2 attachment condition. The speedup at N2 in N1 attachment sentences relative to ambiguous sentences suggests that the parser *sometimes* considers attaching the RC to N2 in ambiguous conditions, but not in N1 attachment conditions.

Importantly, our data did not support the presence of an ambiguity advantage. We hypothesized that such an effect could occur either on N2 or on the spillover position. The posterior probability of an ambiguity advantage of the expected magnitude on N2 was very low (0*.*01), and while it was noteworthy at the spill-over region (0*.*39), its magnitude at the spill-over region would need to compensate for the N1 attachment speedup in order to yield a net ambiguity advantage.

# General Discussion

We conducted two reading experiments to test a prediction made by the strategic underspecification account: If readers strategically underspecify ambiguous sentences, they should do so even when RCs are pre-nominal, as they are in Turkish. Therefore, we expect to see an ambiguity advantage in such cases too. In contrast, if readers don’t underspecify, such as in the SMCM account, we don’t expect to see any ambiguity advantages.

In experiment 1, we tested this prediction in an eye tracking experiment in which a relative clause attachment ambiguity was disambiguated by means of plausibility. We did not find clear evidence for or against an ambiguity advantage in this experiment. While the first-pass reading time and the total fixation time at the critical region did not indicate shorter processing times in ambiguous sentences compared to the preferred N1 attachment conditions, the regression path duration as well as the number of first-pass regressions pointed to a *potential* small slowdown. This finding is consistent with the absence of an ambiguity advantage, as well as with the presence of a relatively small ambiguity advantage.

In experiment 2, we tested the same prediction as in experiment 1, but with a different experiment design. In spite of We found no evidence of

... Here is a recap why we did this experiment, what roughly what the predictions were, what we found (do the experiments even show the same thing?), what it means, and what the caveats are. In a nutshell, we looked for an ambiguity advantage in Turkish and didn’t find one.

This finding is not compatible with a version of underspecification which is capable of suspending RC attachment as it predicts an ambiguity advantage in Turkish, but there is no evidence of one. It is consistent with a race-based model such as SMCM, as well as with a version of underspecification which is incapable of suspending RC attachment. But if it’s the latter, we need to raise the question - why would a parser capable of underspecification, and which carries it out in order to save time be unable to postpone RC attachment in anticipation of an opportunity to save time? Especially if it doesn’t stand to lose any time in the process? Does it only go for certain things, and doesn’t like to ‘gamble?’ ...

# to-do

• Consider stating in every results section that every speed-up or slowdown is relative to the ambiguous condition. Currently there’s too many instances of “compared/relative to the ambiguous condition,” which is good for readers who just scan the text, but it’s also awfully tedious to read.

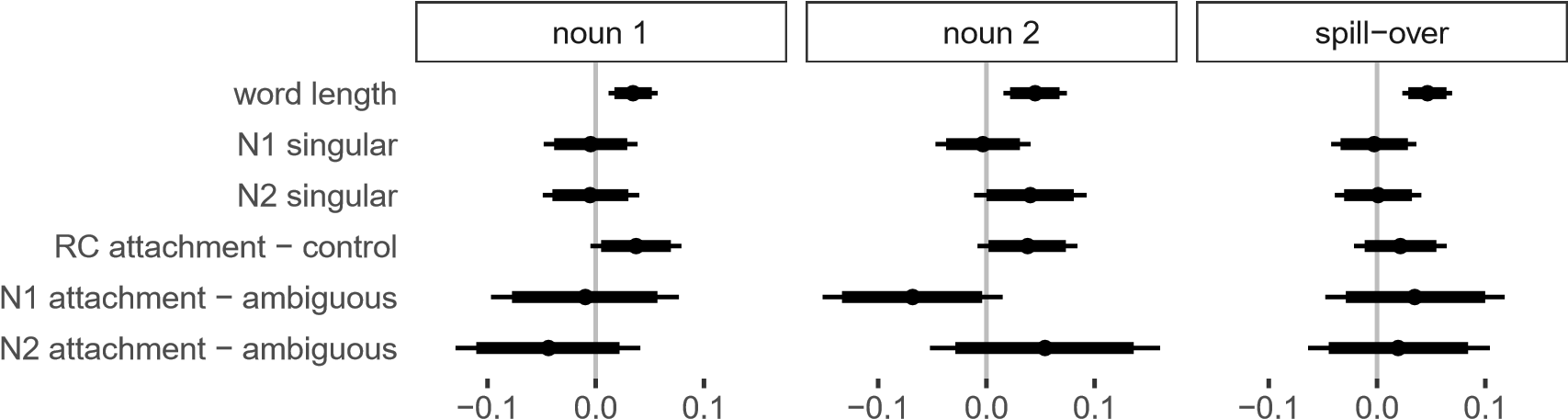
**References**

# Appendix

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to

Loftus and Masson’s method. *Tutorials in Quantitative Methods for Psychology*, *1*(1),

42–45. <https://doi.org/10.20982/tqmp.01.1.p042>



Estimate

# Figure 5

*Estimates and 90% credible intervals (thin lines) as well as 80% credible intervals (thick lines) for the effect (in log-units) of unambiguous RC attachment relative to ambiguous attachment on reading times at the critical regions. The posterior probability that the parameter is smaller than (or larger than) zero is larger than* 0*.*9 *if the 80% credible interval excludes zero, and larger than* 0*.*95 *if the 90% credible interval excludes zero.*

Dempsey, J., & Brehm, L. (2020). Can propositional biases modulate syntactic repair processes? Insights from preceding comprehension questions. *Journal of Cognitive Psychology*, 1–10. <https://doi.org/10.1080/20445911.2020.1803884>

Dillon, B., Andrews, C., Rotello, C. M., & Wagers, M. (2019). A new argument for co-active parses during language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(7), 1271–1286. <https://doi.org/10.1037/xlm0000649>

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel Hierarchical Models*. Cambridge: Cambridge University Press.

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar.* London, New York: Routledge.

Green, M. J., & Mitchell, D. C. (2006). Absence of real evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *55*(1), 1–17. <https://doi.org/10.1016/j.jml.2006.03.003>

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies 2001 - NAACL ’01* (pp. 1–8). Pittsburgh, Pennsylvania: Association for Computational Linguistics. <https://doi.org/10.3115/1073336.1073357>

Kaakinen, J. K., & Hyönä, J. (2010). Task effects on eye movements during reading.

*Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1561– 1566. <https://doi.org/10.1037/a0020693>

Kruschke, J. K. (2015). *Doing Bayesian data analysis, Second Edition: A tutorial with R, JAGS, and Stan*. Burlington, MA: Academic Press/Elsevier.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

<https://doi.org/10.1016/j.cognition.2007.05.006>

Logačev, P., & Vasishth, S. (2013). Em2: A package for computing reading time measures for psycholinguistics.

Logačev, P., & Vasishth, S. (2016). A Multiple-Channel Model of Task-Dependent Ambiguity Resolution in Sentence Comprehension. *Cognitive Science*, *40*(2), 266–298. [https: //doi.org/10.1111/cogs.12228](https://doi.org/10.1111/cogs.12228)

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton, FL: Chapman & Hall/CRC Press.

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau

(2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. [https://doi.org/ 10.20982/tqmp.04.2.p061](https://doi.org/10.20982/tqmp.04.2.p061)

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna,

Austria: R Foundation for Statistical Computing. Retrieved from [https://www.Rproject.org/](https://www.R-project.org/)

Raab, D. H. (1962). Statistical Facilitation of Simple Reaction Times. *Transactions of the*

*New York Academy of Sciences*, *24*(5 Series II), 574–590. [https://doi.org/10.1111/j.21640947.1962.tb01433.x](https://doi.org/10.1111/j.2164-0947.1962.tb01433.x)

Schotter, E. R. (2014). Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading, 27.

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*(1), 201–216. <https://doi.org/10.3758/MC.36.1.201>

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct Attachment Is Not a Form of Lexical Ambiguity Resolution. *Journal of Memory and Language*, *39*(4), 558–592. <https://doi.org/10.1006/jmla.1998.2600>

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and*

*Language*, *52*(2), 284–307. <https://doi.org/10.1016/j.jml.2004.11.003>

van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted Race: A

New Model of Syntactic Ambiguity Resolution. In *Reading as a Perceptual Process* (pp.

621–648). Elsevier. <https://doi.org/10.1016/B978-008043642-5/50029-2>

van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in Sentence Processing: Evidence against Current Constraint-Based and Two-Stage Models. *Journal of Memory and Language*, *45*(2), 225–258. <https://doi.org/10.1006/jmla.2001.2773>

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2019). Bayesian data analysis in the phonetic sciences: A tutorial introduction, 38.

von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1545–1578. [https:](https://doi.org/10.1080/01690965.2012.728232)

[//doi.org/10.1080/01690965.2012.728232](https://doi.org/10.1080/01690965.2012.728232)

Vosse, T., & Kempen, G. (2009). In Defense of Competition During Syntactic Ambiguity

Resolution. *Journal of Psycholinguistic Research*, *38*(1), 1–9. [https://doi.org/10.1007/s1 0936-008-9075-1](https://doi.org/10.1007/s10936-008-9075-1)

Weiss, A. F., Kretzschmar, F., Schlesewsky, M., Bornkessel-Schlesewsky, I., & Staub, A. (2018). Comprehension demands modulate re-reading, but not first-pass reading behavior.

*Quarterly Journal of Experimental Psychology*, *71*(1), 198–210. [https://doi.org/10.1080/ 17470218.2017.1307862](https://doi.org/10.1080/17470218.2017.1307862)

Wotschack, C., & Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY*, 15.

1. This is to avoid any confusion due to the fact that prior literature on RC attachment in Turkish (**Nazik?**; **Ankara1?**; **Ankara2?**) labeled them based on the level of syntactic embedding, with NP2, the noun phrase in the lower syntactic position, being headed by N1. [↑](#footnote-ref-2)