Absence of Evidence for Underspecification in Prenominal Relative Clause Attachment

Pavel Logačev[1], Özgur Aydın[2], and Aylin Müge Tuncer[3]

[1] Boğaziçi University, Istanbul, Turkey
[2] Ankara University, Ankara, Turkey
[3] Muğla Sıtkı Koçman University, Muğla, Turkey

Abstract

Traxler et al. (1998) have found that relative clauses with ambiguous attachment are sometimes read faster than their unambiguous counterparts. Two broad classes of theories account for this phenomenon: Race-based models posit that ambiguous sentences are read faster due to a 'race' between several permissible analyses of the sentence. In contrast, the strategic underspecification account maintains that, under the right conditions, readers underspecify ambiguities in order to save time. We argue that the two accounts make qualitatively different predictions for structures with prenominal relative clauses, such as in Turkish. While the underspecification account predicts an ambiguity advantage in Turkish, race-based accounts predict the absence of such an effect. We present data from two reading experiments in Turkish ($N = 39$ and $N = 184$) in which we find no evidence for a substantial ambiguity advantage in the processing of ambiguous sentences with prenominal relative clauses and argue that this finding poses a major challenge for the strategic underspecification account.

Keywords: relative clause attachment, underspecification, Turkish, sentence processing
Word count: 6,955

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

Correspondence concerning this article should be addressed to Pavel Logačev, E-mail: pavel.logacev@boun.edu.tr

We would like to thank three anonymous reviewers for their helpful comments and suggestions, as well as Shravan Vasishth for extensive comments on earlier versions of this work.

finding: According to their unrestricted race model (URM), the parser uses all available 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 NP1 attachment when it is computed faster than an NP1 attachment, or an NP2 attachment when it is computed faster than an NP1 attachment.

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

b. LOCALLY AMBIGUOUS, NP1 ATTACHMENT
The car of the driver *that had the moustache* was pretty cool.

c. LOCALLY AMBIGUOUS, NP2 ATTACHMENT
The driver of the car *that had the moustache* was pretty cool.

Thus, according to the URM, the local ambiguity is disambiguated in all attachment conditions *before* any disambiguating information is processed. Sometimes it is resolved towards one reading and sometimes towards the other. In consequence, the 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. This absence of reanalysis surfaces as an ambiguity advantage. The ambiguity advantage effect has been replicated in multiple experiments (Dillon, Andrews, Rotello, & Wagers, 2019; van Gompel, Pickering, Pearson, & Liversedge, 2005; van Gompel, Pickering, & Traxler, 2001; von der Malsburg & Vasishth, 2013).

Multiple alternative explanations of the ambiguity advantage have been proposed, such as surprisal (Hale, 2001; Levy, 2008), and specific implementations of constraint-based models (Green & Mitchell, 2006; Vosse & Kempen, 2009). For example, Levy's (2008) independently motivated surprisal account 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 this 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 shorter reading times in globally ambiguous conditions compared to the locally ambiguous conditions are a result of the lower surprisal in ambiguous sentences.

Another influential account of the ambiguity advantage 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. It posits that because participants in the studies by Traxler et al., as well as Van Gompel et al. were asked fairly superficial comprehension questions, they decided that ambiguous sentences do not need to be disambiguated as they did not expect to be asked about the ambiguous grammatical relation. In other words, readers underspecify RC attachment when they believe that it is task-irrelevant. For example, RCs with ambiguous attachment such as in sentence (1a) are not attached to NP1 or NP2, but instead are loosely associated with the entire complex NP indicating attachment to any constituent within it. Swets et al. further stipulate that underspecification of ambiguous relative clause attachment saves time, which explains the ambiguity advantage attested in the above-mentioned studies. Importantly, underspecification is assumed to be a viable strategy only when task demands do not require disambiguation. Swets et al. tested this explanation in 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 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 can thus switch underspecification on and off.

(2) a. GLOBALLY AMBIGUOUS
The maid of the princess who scratched herself in public was terribly humiliated.

b. LOCALLY AMBIGUOUS, NP1 ATTACHMENT
The son of the princess who scratched himself in public was terribly humiliated.

c. LOCALLY AMBIGUOUS, NP2 ATTACHMENT
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 for two reasons: Firstly, a recent reanalysis of the Swets et al. data (Vasishth, 2021) shows that the key finding, an interaction between question type and the effect of RC attachment is significant *only* under an analysis of raw reading times with an ANOVA: In an ANOVA analysis of log-reading times, or an analysis with linear mixed effects models (raw or log-transformed) the interaction fails to reach significance. In light of the fact that the combination of F1 & F2 ANOVAs has an inflated Type I error rate (Baayen, Davidson, & Bates, 2008), and that the analysis of untransformed reaction times is more likely to be affected by potential outliers, this finding indicates that the Swets et al. data does not provide unambiguous evidence for strategic underspecification. Secondly, Logačev and Vasishth (2016) have demonstrated by means of simulations that Swets et al.'s findings are compatible with some race-based accounts under the assumption that relative clauses are attached only after they have been processed to a sufficient degree, which may also be a necessary assumption of the underspecification account. According to their proposal, the *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 time 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 in general (Dempsey & Brehm, 2020; Kaakinen & Hyönä, 2010; Schotter, Bicknell, Howard, Levy, & Rayner, 2014; Weiss, Kretzschmar, Schlesewsky, Bornkessel-Schlesewsky, & Staub, 2018; Wotschack & Kliegl, 2013).

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 any RC attachment-related operations 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 postnominal relative clauses, such as in English in (1) and (2), their predictions diverge for prenominal relative clauses, such as in Turkish.

Ambiguity resolution in pre- and postnominal 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 postnominal, i.e., they follow the noun phrases

they modify. As a result, when readers have processed a postnominal 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 prenominal relative clauses, such as Turkish (Göksel & Kerslake, 2005).

In constructions like (3), the potential attachment sites form a complex noun phrase with the structure *[NP1-genitive [N2-possessive]],* in which the 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 lives in the city center]'* modifies either the complex subject noun phrase headed by *'professor'* (NP2), or its possessor *'secretary'* (NP1). Thus, this sentence can either mean that the professor lives in the city center (local attachment; to the NP headed by N1), or that the professor's secretary lives in the city center (non-local attachment; to the NP headed by N2). For the sake of clarity, we will refer to these attachment conditions as NP1 and NP2 attachment, in reference to the position of the NP's head noun.

Prior research suggests that sentences like (3) display an NP1 attachment preference: In a questionnaire study with sentences like (3), Kırkıcı (2004) found a weak tendency towards NP1 attachment (55%-63%). Similarly, Dinçtopal-Deniz (2010) found an NP1 attachment preference of 66%.

(3) Şoför, [şehir merkezinde oturan]$_{RC}$ profesörün sekreterini gördü.
driver, in the city center living professor's secretary saw
'The driver saw the secretary of the professor who was living in the city center.'

The earliest point in this sentence at which an attachment can be made is the first noun (*'professor'*). 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 *'professor'* and *'who lives in the city center'* can be established. At this point, 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 to be the possessor of a complex noun phrase and that the parser can thus expect another potential attachment site at the next word.

The fact that the two possible attachment sites become available *successively* in structures like (3) makes them particularly interesting for testing the idea that the parser strategically underspecifies relative clause attachment in order to save time. This is because according to the strategic underspecification account, the parser should underspecify RC attachment when the task demands allow, thus yielding an ambiguity advantage, while models like the SMCM (and URM) predict no such effect since the relatively late availability of the second attachment site would significantly delay any attempt to attach the relative clause to it.

In order to test for the presence of an ambiguity advantage in Turkish, we created three types of sentences similar to (4a), (4b), and (4c). RC attachment was controlled by means of the plausibility: For example, in sentence (4b) the relative clause *"who was shouting in the school building"* can attach only to *principal* (NP1), but not to *voice* (NP2), whereas in sentence (4c) the relative clause *'who was fired'* can attach only to *cook* (NP2), but not to *yacht* (NP1). 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, NP1 ATTACHMENT

[Okulda bağıran]$_{RC}$ müdürün sesi caddenin karşısında duyuldu.
in the school building shouting principal voice street.GEN opposite hear.
'The voice of the principal who was shouting in the school building was audible from across the street.'

c. UNAMBIGUOUS, NP2 ATTACHMENT

[İşten çıkarılan]$_{RC}$ yatın aşçısı işverenden parasını almadı.
from work dismissed yacht.GEN cook from employer his money did not take
'The yacht's cook who was fired by his employer did not take his money.'

Under superficial task demands, *early underspecification* predicts that the parser should underspecify RC attachment in NP1 attachment and ambiguous conditions during the processing of NP1, since it routinely underspecifies in order to reduce the processing load. Because the underspecified attachment is disambiguated at N2 in NP1 attachment conditions such as (4b), ambiguous sentences should be processed faster than NP1 attachment sentences at N2. No difference in reading time between ambiguous and NP1 attachment conditions is expected at N1, because initial underspecification is carried out in all conditions.

The *late underspecification* account assumes that the parser does not preemptively underspecify in anticipation of a possible ambiguity, and that it delays attachment-related operations until it can unambiguously determine how many attachment sites are available. As a result, it also predicts an ambiguity advantage, but for slightly different reasons. According to this account, when the parser is at N1, it will delay RC attachment until it can determine whether the structure is ambiguous, which happens at or following N2. Once this has been determined, the parser should underspecify RC attachment in ambiguous attachment conditions, and attach the RC unambiguously in unambiguous attachment conditions. The consequence should be an ambiguity advantage at N2 or the spill-over region.

The SMCM, on the other hand, posits that the parser should compute one attachment as soon as it can when task demands do not require the computation of both interpretations. Because the first attachment site becomes available at N1 and because there is no mechanism by which RC attachment can be delayed, the parser will always attempt to carry out RC attachment while at N1. This process can be successfully carried out in the NP1 attachment and in the ambiguous attachment conditions (4b) and (4a). Subsequently, at N2, no additional RC attachment attempts take place because the occasional comprehension questions in both of our experiments are not expected to encourage the computation of a second interpretation of the sentence. In NP2 attachment sentences

such as (4c), in contrast, the RC attachment attempt at N1 will fail, possibly resulting in slowed reading relative to the other two conditions. Subsequently, at N2, RC attachment can be carried out successfully, resulting in a slowdown relative to the other two conditions, where it is even attempted since it has successfully been carried out at N1. As a result, the SMCM does not predict an ambiguity advantage in Turkish sentences like (4a) because arguably NP1 attachment is completed before the head of NP2 is even processed.

The most likely prediction of the surprisal account (Levy, 2008) is an ambiguity advantage at N1, as this is where the N1 attachment condition is disambiguated, thus leading to higher surprisal in N1 attachment conditions compared to ambiguous conditions. However, we cannot be certain of this, as surprisal's predictions depend on a number of underlying assumptions regarding the grammar of Turkish as reflected in the language model used in surprisal calculations. The situation is similar for constraint-based models (Green & Mitchell, 2006; Vosse & Kempen, 2009) as its predictions depend on a set of specific constraints assumed in the processing of this structure, as well as their timing.

In sum, the underspecification account predicts an ambiguity advantage in sentences like (4), while the SMCM predicts the absence of such an effect. We conducted one eye-tracking experiment and one self-paced reading experiment to test these predictions. In both experiments, task demands were kept superficial. In experiment 1 ($N = 39$), we disambiguated RC attachment by means of animacy, while in experiment 2 ($N = 184$), RC attachment was disambiguated morphosyntactically.

Importantly, the SMCM predicts no significant attachment-related speed-up in ambiguous conditions, which corresponds to a statistical null result. Therefore, as part of our analysis, we will use Bayes Factors (e.g., Lee & Wagenmakers, 2014; Nicenboim & Vasishth, 2021; Schönbrodt & Wagenmakers, 2018) to quantify the relative evidence for the hypotheses that the data are compatible with (a) the absence of an ambiguity advantage, and (b) the presence of an ambiguity advantage of a previously observed magnitude. Thus, we operationalized an ambiguity advantage as the observation of a speed-up in the ambiguous condition relative to the preferred NP1 attachment conditions and its magnitude as the difference between the fastest unambiguous condition and the ambiguous condition. Table 1 lists previously observed instances of the ambiguity advantage (Swets et al., 2008; Traxler et al., 1998; van Gompel et al., 2005, 2001). The median magnitude of the ambiguity advantage effect amounted to a $46ms$ difference between the ambiguous condition and the preferred unambiguous condition (SD: $30\,ms$, range: $22\,ms - 120\,ms$). This effect size corresponds to 7% of the average reading time at that region (SD: 5%, range: $3\% - 19\%$). Thus, if an ambiguity advantage is observed, we expected it to be of approximately that magnitude.

Experiment 1

We conducted experiment 1 with the aim of testing for the presence of an ambiguity advantage at the second noun (N2) or the spill-over region. The presence of an ambiguity advantage would be compatible with both types of underspecification, but not with the SMCM (or the URM). In order to create task demands which encourage underspecification, participants were asked comprehension questions occasionally, most of which were superficial. To discourage participants from skipping over the relative clause, we asked five

Table 1

Effect sizes of previously observed instances of the ambiguity advantage.

article	experiment	position	measure	effect size (ms)	effect size (%)
Traxler et al. (1998)	Experiment 1	disambiguating region	TFT	82	15
	Experiment 2	disambiguating region	TFT	27	7
Von Gompel et al. (2001)	Experiment 1	spill-over region	TFT	22	3
	Experiment 2	spill-over region	RPD	120	19
		spill-over region	TFT	41	6
Von Gompel et al. (2005)	Experiment 1	disambiguating region	RPD	26	6
		disambiguating region	TFT	50	11
		spill-over region	RPD	86	12
		spill-over region	TFT	79	10
	Experiment 2	disambiguating region	TFT	36	6
		spill-over region	RPD	31	4
		spill-over region	TFT	83	10
Swets et al. (2008)	superficial questions	spill-over region	SPR	52	5
	occasional questions	spill-over region	SPR	36	3

Table 2

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	spill-over
ambiguous	7.4 (1.2; 6-10)	5.8 (1.1; 4-7)	5.4 (1.0; 4-7)	5.5 (2.0; 3-9)
N1 attachment	6.7 (0.9; 5-8)	5.7 (1.2; 4-8)	5.3 (1.2; 3-7)	5.5 (1.7; 3-8)
N2 attachment	7.8 (1.1; 7-10)	5.7 (1.2; 3-7)	5.3 (1.2; 3-7)	5.7 (3.1; 3-13)

RC-related comprehension questions in unambiguous attachment conditions over the course of the experiment.

Method

Materials

We constructed a total of 36 experimental sentences such as in (4), i.e., twelve sentences for each attachment condition. Disambiguation was effected by means of animacy and plausibility. Relative clauses always comprised two words. The critical region and the spillover region were approximately equal in length across the three conditions (cf. Table 2).

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. 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 experiments took place at Ankara University Linguistics Laboratory, 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. The degree of angle per character was 0.074 cm (Courier New, 36 point). During the performance of the sentence processing task, the eye-tracker device (RED 500 by SensoMotoric Instruments, SMI) with a chin rest recorded the participants' binocular eye movements at a sampling rate of 500Hz, 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. The practice session began with a 5-point calibration, followed by a 4-point validation procedure. The calibration procedure was repeated when necessary or when the validation resulted in an average error of more than 0.5 degrees.

During the testing phase, each participant read 48 sentences (36 test sentences and 12 filler sentences) in two sentence blocks of 24 sentences each. Before each sentence, a fixation cross appeared for 500 ms trigger duration, in order to help the participants to fixate their eyes on 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. Comprehension questions about the sentence were asked on every second trial: 19 of the 24 questions were superficial, and 5 questions were about the RC attachment ambiguity. All RC questions were asked in unambiguous attachment conditions. Participants answered questions by selecting one of two options 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 (R Core Team, 2018) for computing reading time measures from the eye movement record. We used the *tidyverse* and *ggplot2* packages (Wickham, 2016, 2017) for data processing and plotting, and the R packages *brms* (Bürkner, 2017) 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, 2018) to the eye-tracking measures of interest. Reading-time measures were modeled using generalized hierarchical linear models assuming log-normally distributed residuals. Regression probability was modeled with generalized hierarchical linear models with a logit link function.

All models were fitted using four chains with 2,000 warmup samples and 5,000 post-warmup samples. We used a normal prior ($N(\mu = 5.75, \sigma = 1)$) for the intercept of the log-normal models of reading time and the symmetrical prior $N(0, 0.2)$ for all slope coefficients. In the logit model of regression probability, $N(0, 3)$ was used as the intercept prior, while all slope coefficients received the prior $N(0, 1)$. All models included fixed effects of attachment coded using centered simple contrasts comparing each of the unambiguous attachment conditions to the ambiguous condition. Centered log 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 on the 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 for more easily interpretable effect sizes.

$$BF_{01} = \frac{p(data|M_0)}{p(data|M_1)} \tag{1}$$

We calculated Bayes Factors quantifying the evidence against the presence of an ambiguity advantage for a range of priors (e.g., Lee & Wagenmakers, 2014) using the Savage-Dickey method implemented in *brms* (Bürkner, 2017). The Bayes Factor BF_{01} for two models M_0 and M_1 quantifies the amount of evidence the data provides for model M_0 relative to M_1. It represents the ratio of posterior odds of the two models, after seeing the data, to their prior odds. Assuming equal prior odds for models M_0 and M_1, BF_{01} corresponds to the ratio of the marginal likelihoods of the data, given the models, as in equation 1.

Importantly, Bayes Factors depend on the prior distribution of the parameter of interest, and thus require a specific hypothesis about the size of the effect in order to quantify the evidence for its absence. As table 3 shows, the prior we used for the estimation of slope coefficients ($N(0, 0.2)$) is rather wide, as it assigns a probability of 0.17 to effect sizes of more than 19%.[1] This means, for instance, that the prior stipulates that when the average reading time is 400 ms the effect is quite likely to be larger than 75 ms. Since unrealistically wide priors are likely to decrease the marginal likelihood $p(data|M)$, they may put the models being compared on unequal footing. Therefore, we computed Bayes Factors for a range of narrower priors for *the difference between the preferred N1 attachment condition and the ambiguous condition* which correspond to different assumptions about hypothesized effect sizes. All priors are listed in table 3. All priors except $\mu = 0, \sigma = 0.2$ allocate a small-to-negligible amount of probability mass to effect sizes larger than 19% but differ in the amount of probability mass allocated to effect sizes which are substantially smaller than previously observed instances of the ambiguity advantage (smaller than 3%). All priors except $\sigma = 0.02$ allocate a sizeable amount of probability mass to the range $3\% - 19\%$ which is compatible with previously observed instances of the ambiguity advantage.

[1] The specification of an effect size as a percentage of the intercept is a consequence of the fact that additive effects on the log-scale are multiplicative in their original units.

Table 3

Allocation of probability mass to effect sizes for different prior widths.

prior width	less than 3%	between 3% and 19% (previously observed instances of the ambiguity advantage)	more than 19%
$\mu = 0, \sigma = 0.2$	.56	.27	.17
$\mu = 0, \sigma = 0.1$	.62	.35	.03
$\mu = 0, \sigma = 0.08$	.65	.35	.01
$\mu = 0, \sigma = 0.06$	.69	.31	$< .01$
$\mu = 0, \sigma = 0.04$	.77	.23	$< .01$
$\mu = 0, \sigma = 0.02$	.93	.07	$< .01$

Data Availability

All data and R code used in the data analysis is available at `https://osf.io/r8dm7/`.

Results

The average comprehension questions accuracy was 78.8%. Figure 1 show a summary of the stimated marginal means and 95% credible intervals of the eye-tracking measures computed for the the critical regions using the `emmeans` package (Lenth, 2021): (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 95% and 90% credible interval for each position (noun 1, noun 2, spill-over region).[2] The estimates indicate a slowdown in NP2 attachment conditions in first-pass reading times (*N2:* CrI [-13; 53], $P(\beta > 0) = .89$; *spill-over:* CrI [-17; 49], $P(\beta > 0) = .84$), with stronger evidence for a slowdown in regression-path duration (*N2:* CrI [8; 119], $P(\beta > 0) = .99$, *spill-over:* CrI [-12; 81], $P(\beta > 0) = .93$). In total fixation time, we found a slowdown in NP2 attachment conditions at N2 (CrI [-13; 117], $P(\beta > 0) = .94$), and an indication of an effect in the same direction at the spill-over region (CrI [-31; 77], $P(\beta > 0) = .81$). We further found an NP2 attachment-related slowdown in regression probability at N2 (CrI [-13.2; 117.3], $P(\beta > 0) = .94$).

[2]The posterior probability that the parameter is smaller than (or larger than) zero is larger than 97.5% if the 95% credible interval excludes zero and larger than 95% if the 90% credible interval excludes zero.

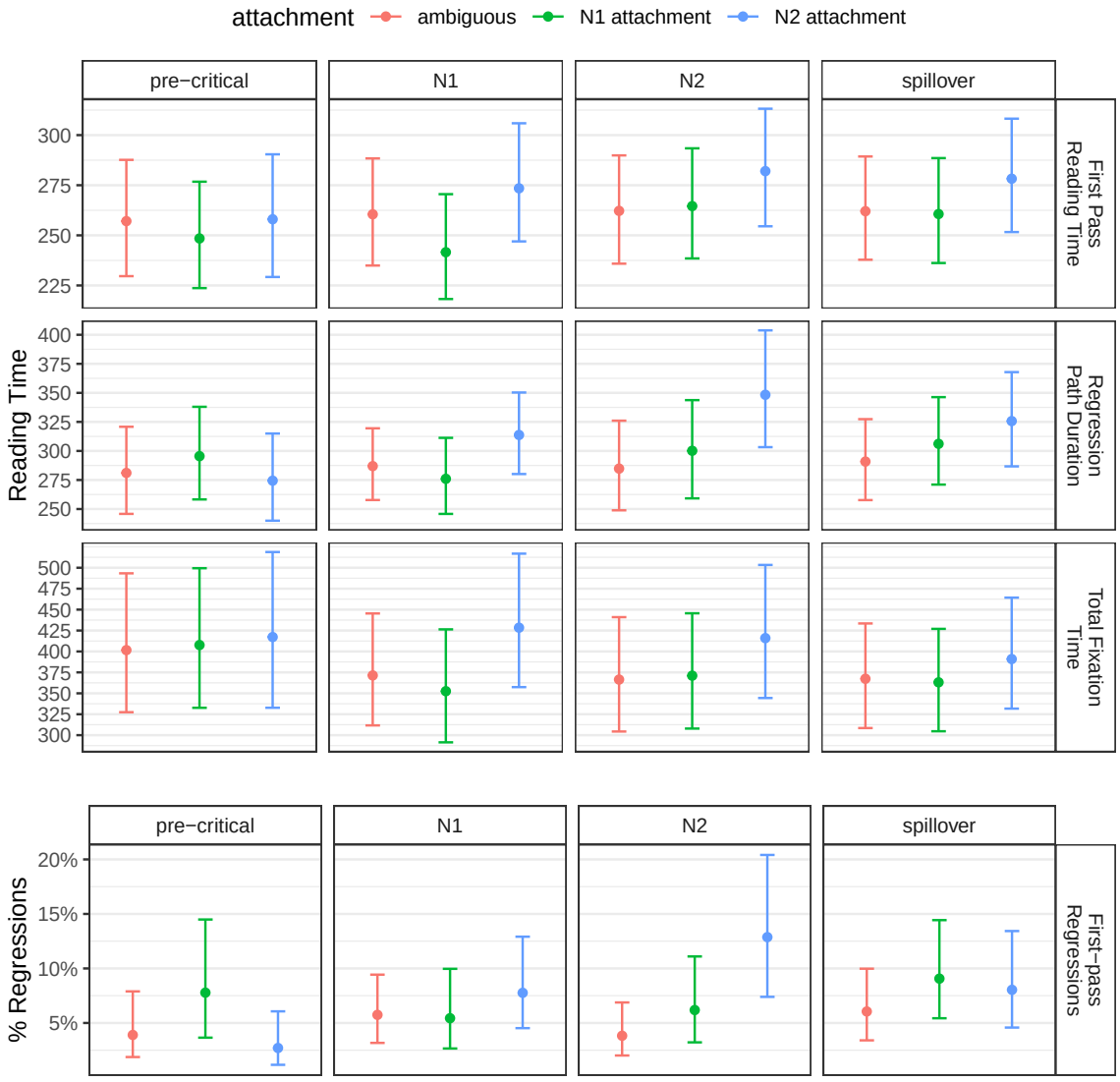


Figure 1

Estimated marginal means and 95% credible intervals of first-pass regression probabilities and back-transformed log-reading time measures FPRT, RPD, and TFT for the critical regions by condition. All estimates and credible intervals were obtained using the `emmeans` *package (Lenth, 2021).*

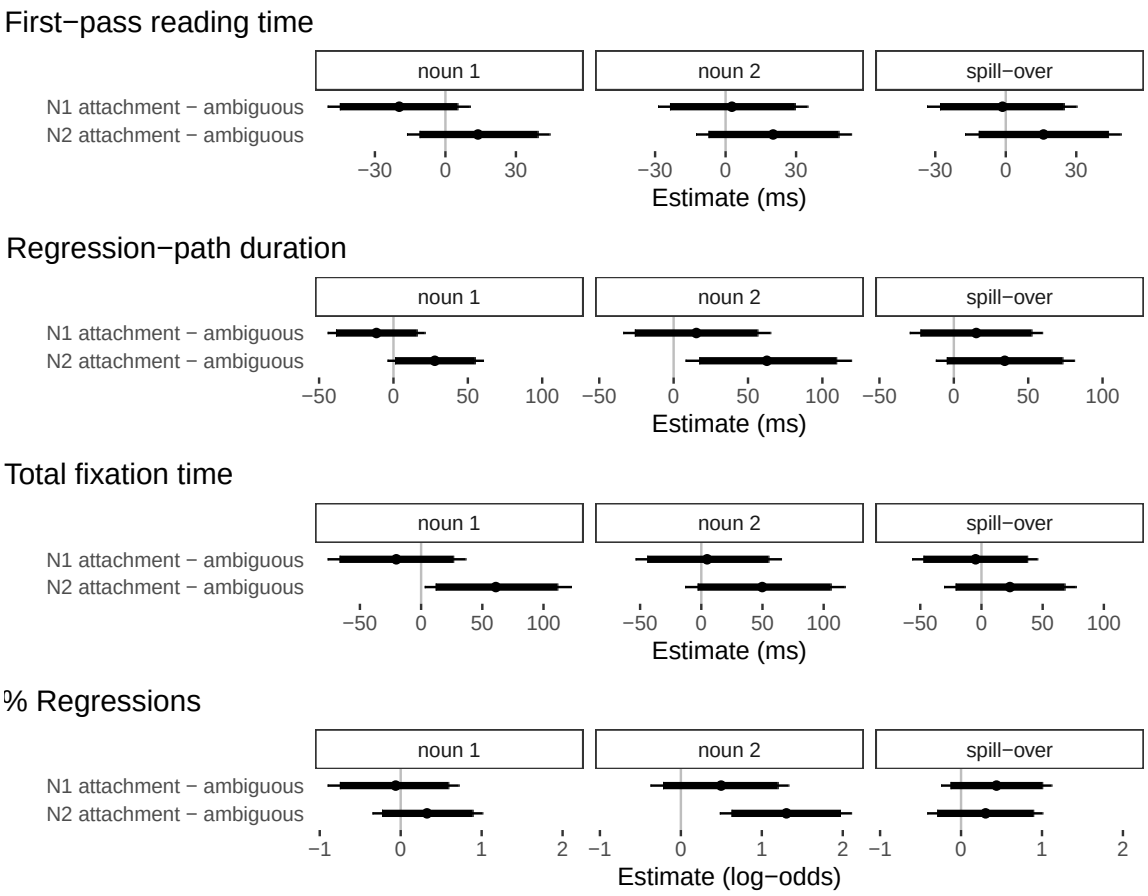


Figure 2

Estimates and 95% *credible intervals (thin lines) as well as* 90% *credible intervals (thick lines) for the analyses of first-pass reading time, regression-path duration, total fixation time, and regression probability.*

While most eye tracking measures at most regions reflected an NP2 attachment-related slowdown, only regression path duration and regression probability suggested a slowdown in the NP1 attachment conditions. Specifically, in RPD at N2 (CrI [-34; 65], P($\beta > 0$) = .74) and at the spill-over region (CrI [-30; 59], P($\beta > 0$) = .76), as well as regression probability at N2 (CrI [-0.4; 1.3], P($\beta > 0$) = .88).

Figure 3 shows the Bayes Factors in favor of the hypothesis of no difference between NP1 attachment conditions and ambiguous conditions. While the Bayes Factor indicates a small amount of evidence for the hypothesis of *no difference* between ambiguous and N1 attachment conditions for all but the most narrow prior ($\sigma = 0.02$), it fails to provide unambiguous evidence in favor of that hypothesis.

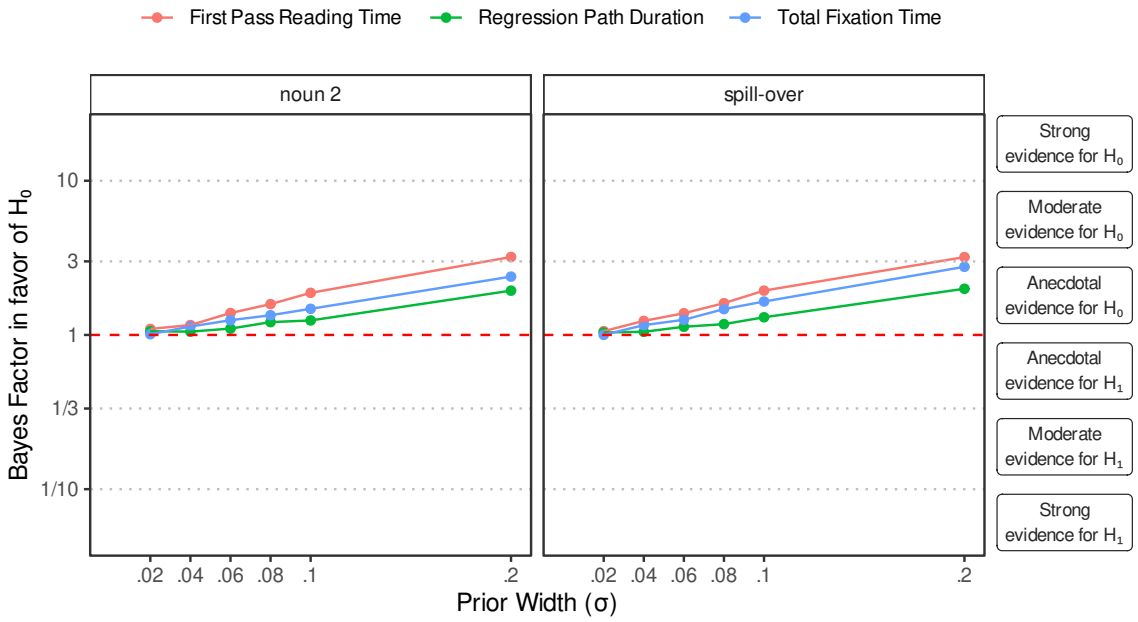


Figure 3

Bayes Factors in favor of the absence of a difference in reading times between the ambiguous attachment condition and the N1 attachment condition, at two positions in the sentence: noun 2 and the spill-over region. Classification of evidence strength following Jeffreys (1961) as cited in Lee and Wagenmakers (2014).

Discussion

We found evidence for substantial processing difficulty in unambiguous NP2 attachment sentences relative to the ambiguous attachment condition, while the effect of unambiguous NP1 attachment was less clear. In the NP2 attachment condition, we found an indication of processing difficulty in most eye-tracking measures starting with regression path duration on N1, suggesting that reading in NP2 attachment sentences is initially slowed by the the fact that the RC cannot be attached to the first structurally available noun phrase, headed by N1. The subsequent slowdown relative to the ambiguous baseline may be due to the time the parser spends in carrying out RC attachment to the noun phrase headed by N2.

The results concerning the relative difficulty of the NP1 attachment sentences are less clear-cut: We found no appreciable slowdown in first-pass reading time or total fixation time at N2 in the NP1 attachment condition relative to the ambiguous condition. However, the regression path duration as well as the regression probability indicated somewhat higher processing difficulty in the NP1 attachment conditions compared to the ambiguous conditions. While the Bayes Factor analysis indicated weak evidence in favor of an absence of an ambiguity advantage, it failed to provide unambiguous evidence in favor of this hypothesis.

Experiments 2A & 2B

The second experiment was designed to address three issues which may complicate the interpretation of the first experiment: Firstly, although participants in Experiment 1 were asked only 5 comprehension questions about relative clause attachment, and only in unambiguous conditions, their presence may still have discouraged underspecification. Secondly, the analysis of multiple eye tracking measures at multiple regions increases the chances that one of them exhibits a theoretically interesting effect purely by chance, even in the absence of such an effect (e.g., von der Malsburg & Angele, 2017). Lastly, a plausibility-based disambiguation of RC attachment may have caused between-participant or between-item variability in the time course of disambiguation, thus making its effect more difficult to detect.

In order to address these potential shortcomings, we conducted a self-paced reading experiment in which we asked only occasional superficial questions on one third of the trials, and used a different stimulus design. 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 the complex noun phrase *'fan(s) of the football player(s)'*. 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: NP1 attachment, NP2 attachment and ambiguous attachment, as illustrated in (6). Because plurals were formed through the addition of the plural suffix *-ler/-lar*, reading times for plurals and singulars may vary due to word length. 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 NP1. In the examples in (7) it is *'Fenerbahçeli'* (meaning *'of Fenerbahçe'*, a well-known Turkish football team). These additional conditions served as a baseline, and any differences between experimental conditions in (6) and control conditions in (7) may be considered effects of RC attachment.

As in Experiment 1, both underspecification accounts predict a speed-up in the ambiguous condition at the second noun, compared to the NP1 attachment condition. This prediction translates into an interaction between modifier type and the grammatical number of the second noun. That is, NP1 attachment sentences should exhibit a larger slowdown relative to their controls than the ambiguous condition relative to its control. In contrast, the SMCM as well as the late underspecification account, predict the absence of such an interaction. It predicts the absence of an ambiguity advantage, which means that there should not be any significant interaction between modifier type and the grammatical number

of the second noun in the RC attachment conditions.

(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. NP1 ATTACHMENT (PLURAL-SINGULAR)
[birbirini döven]$_{RC}$ futbolcu-lar-ın hayran-ı
each other hit football player-PL-GEN fan.SG-POSS

c. NP2 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. NP1 ATTACHMENT CONTROL (PLURAL-SINGULAR)
[Fenerbahçe-li]$_{ADJ}$ futbolcu-lar-ın hayran-ı
Fenerbahce-ADJ hit football player-PL-GEN fan.SG-POSS

c. NP2 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

The experiment was conducted in the lab (Experiment 2A) and online (Experiment 2B) using the same stimuli. Participants in both experiments provided informed consent and the procedures in this study were compliant with the ethical principles outlined in the Helsinki Declaration on research involving human subjects.

Experiment 2A. Thirty-six students of Anadolu University in Eskişehir, Turkey participated in the experiment. All participants were native speakers of Turkish, with an age range of 19-29 years.

Experiment 2B. A total of 148 students of four universities in Turkey participated in the experiment on-line in exchange for course credit (Boğaziçi University: 116, Ankara University: 13, Muğla Sıtkı Koçman University: 13, İstanbul Medipol Üniversitesi: 6). All participants were native speakers of Turkish, with an age range of 17-47 years (mean: 22 years).

Procedure

The task was self-paced non-cumulative word-by-word reading. 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'` or `'no'` by pressing the corresponding button on the keyboard.

Experiment 2A. Presentation and recording was done in a quiet room, using the Linger software package, version 2.94 by Doug Rohde. One experimental session took approximately 40 minutes to complete. Participants were paid 15 Turkish Lira for participation.

Experiment 2B. The experiment was run online, using the web-based platform PCI Farm (Zehr & Schwarz, 2018). One experimental session took approximately 30 minutes to complete.

Statistical Analysis

Because the experimental items were identical in Experiments 2A and 2B, we pooled the data from both experiments for the purposes of the statistical analysis in order to increase statistical power for detecting an ambiguity advantage effect.[3] We excluded the data of seven participants from Experiment 2B from further analysis because they failed to answer more than 25% of the comprehension questions correctly. We also excluded all trials in which the reading time on one of the words between the pre-critical region and the

[3]We did not include a predictor for experiment modality (indicating *'online'* or *'in the lab'*) because potential interactions between the magnitude of a potential ambiguity advantage in Turkish and experiment modality are beyond the scope of this paper.

spill-over region was shorter than $150\,ms$ or longer than $3000\,ms$, resulting in exclusion of further 2.2% of the data (170 out of $7,644$ trials).

The data analysis followed the same logic as in Experiment 1: All reading time measures were modeled using Bayesian generalized hierarchical linear models assuming log-normally distributed residuals We used a normal prior ($N(\mu = 6, \sigma = 1)$) for all model intercepts, and a symmetrical prior for all slope coefficients ($N(0, 0.2)$). All models were fitted using four chains with $2,000$ warmup samples $5,000$ post-warmup samples each.

All models included fixed effects of (i) centered log-word length, (ii) the availability of NP1 and N2 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 NP1 and NP2 for RC attachment. Log 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 and correlations between random intercepts and slopes.

The grammatical number of N1 and N2 which determined the potential availability of NP1 and NP2 for RC attachment when an RC was present was coded using centered contrasts. As a result, the relative difficulty of RC attachment in the three attachment conditions corresponded to the interactions between N1 and N2 number and the presence of a relative clause: (i) The interaction between the presence of an RC and N1 number corresponded to the difference between the effect of N1 number in the RC conditions ((6b) vs. (6a)) and its effect in the control conditions ((7b) vs. (7a)), that is, *effect of NP1 attachment - effect of ambiguous attachment*. (ii) Similarly, the interaction between the presence of an RC and N2 number corresponded to the difference between the effect of N2 number in the RC conditions ((6c) vs. (6a)) and its effect in the control conditions ((7c) vs. (7a)), that is *effect of NP2 attachment - effect of ambiguous attachment*. In sum, because the coefficients associated with the interaction contrasts represent the differences in the effect of NP1 and NP2 availability associated with the presence of a relative clause, they can be interpreted as the respective effects of unambiguous NP1 and NP2 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*).

As in Experiment 1, we used the Savage-Dickey method to calculate Bayes Factors quantifying the evidence against the presence of an ambiguity advantage for a range of priors. Because in this experiment, the critical parameter was the interaction between the presence of an RC and N1 number, we fitted all models with all priors in table 3.

Data Availability

All data and R code used in the data analysis is available at `https://osf.io/r8dm7/`.

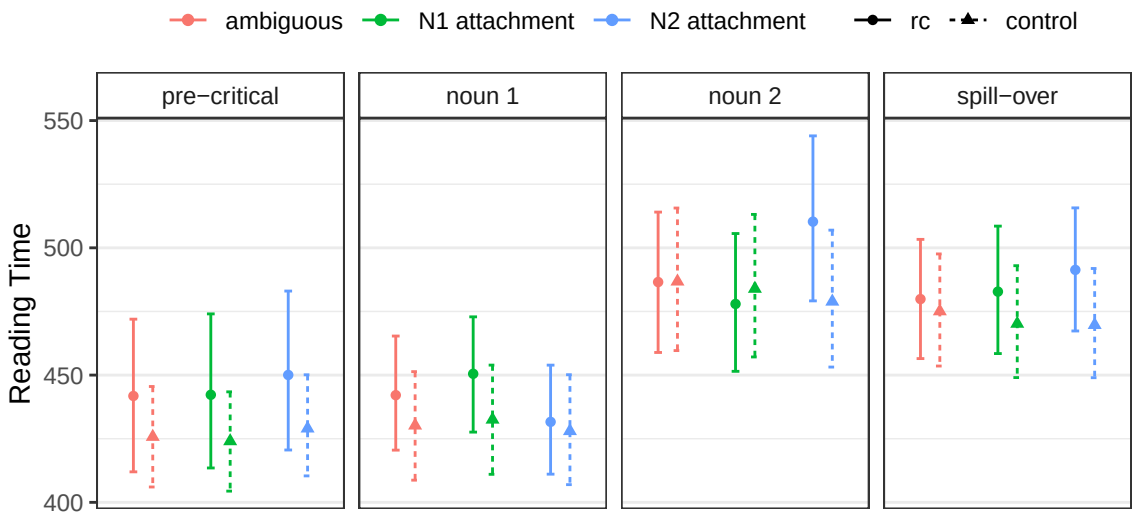


Figure 4

Estimated marginal means and 95% credible intervals of back-transformed log-reading time for the critical regions by condition. All estimates and credible intervals were obtained using the `emmeans` *package (Lenth, 2021).*

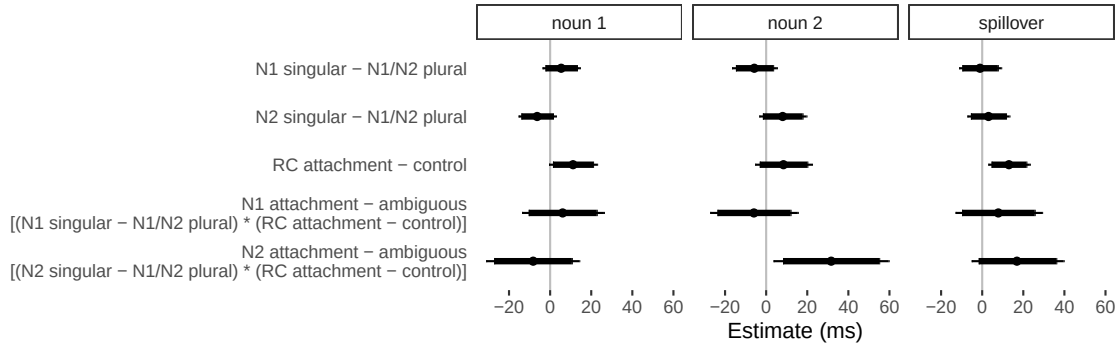


Figure 5

Estimates and 95% credible intervals (thin lines) as well as 90% credible intervals (thick lines) for the effect (in milliseconds) of unambiguous RC attachment relative to ambiguous attachment on reading times at the critical regions.

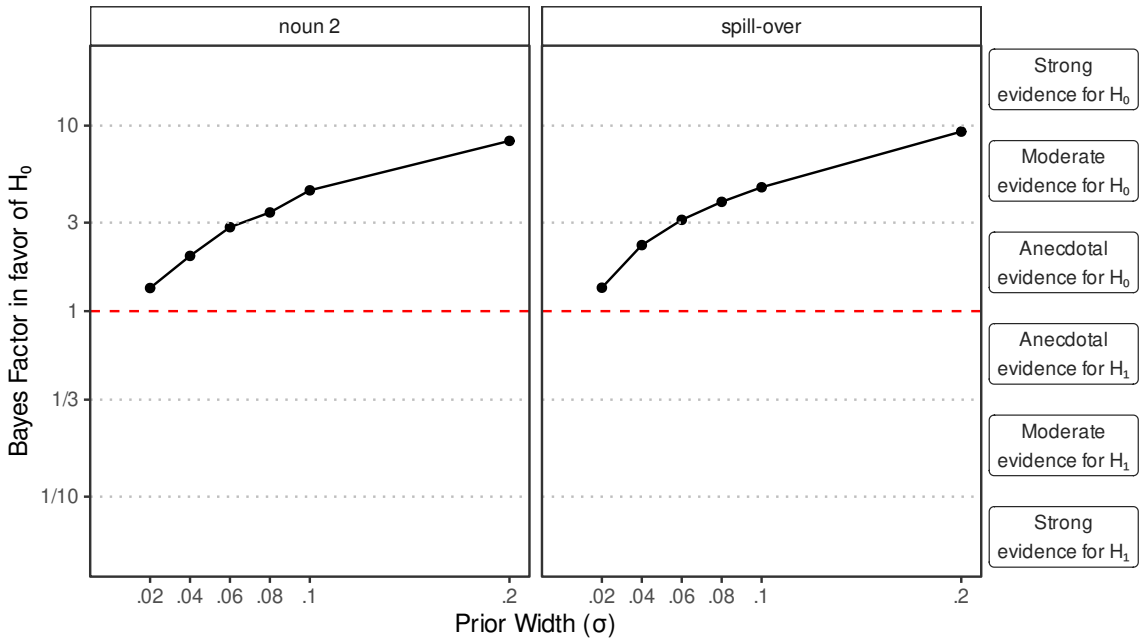


Figure 6

Bayes Factors in favor of the absence of a difference in reading times between the ambiguous attachment condition and the N1 attachment condition, at two positions in the sentence: noun 2 and the spill-over region. Classification of evidence strength following Jeffreys (1961) as cited in Lee and Wagenmakers (2014).

Results

Figure 4 shows the average log-reading times transformed back to the original scale (milliseconds). Figure 5 shows the estimates and 95% and 90% credible intervals for the key contrasts at the critical regions in milliseconds. The effects of RC attachment correspond to the interactions between the presence of a relative clause and the contrasts encoding the grammatical number of N1 and N2.[4]

As figure 5 shows, we found that N1 and N2 were read faster in their singular form than in in their plural forms (*N1:* CrI [-4; 15], P($\beta > 0$) = .88, *N2:* CrI [-15; 3], P($\beta > 0$) = .09). Moreover, it shows an across-the-board slowdown in reading in RC attachment conditions compared to control conditions (*N1:* CrI [-1; 23], P($\beta > 0$) = .97, *N2:* CrI [-5; 22], P($\beta > 0$) = .89, *spill-over:* CrI [3; 23], P($\beta > 0$) = .995).

We found no indication of differences in reading time between the *NP1 attachment condition* and the ambiguous attachment condition. However, we observed a slowdown in the *NP2 attachment condition* relative to the ambiguous condition at N2 (CrI [4; 60], P($\beta > 0$) = .99) and at the spill-over region (CrI [-5; 40], P($\beta > 0$) = .93).

The posterior probability that the ambiguity advantage effect (the difference between the NP1 attachment and the ambiguous condition) was larger than the small-

[4]The posterior probability that the parameter is smaller than (or larger than) zero is larger than 97.5% if the 95% credible interval excludes zero and larger than 95% if the 90% credible interval excludes zero.

est ambiguity advantage attested in the literature (3%) was ($P(log\ \beta \geq 0.03) = 0.03$, $P(\beta \geq 22\,ms) = 0.01$) at N2, and ($P(log\ \beta \geq 0.03) = 0.10$, $P(\beta \geq 22\,ms) = 0.09$) at the spill-over region.

Figure 6 shows Bayes Factors for the absence of a difference between ambiguous and N1 attachment conditions. The analysis shows a moderate or near-moderate amount of evidence against models with $\sigma = 0.2/0.1/0.08/0.06$: All BFs were > 3 with the exception of the BF for $\sigma = 0.06$ at the N2 region, which was $BF_{01} = 2.8$. Meanwhile, the evidence against models with $\sigma = 0.04$ is relatively weak (*N2:* $BF_{01} = 2$, *spill-over:* $BF_{01} = 2.3$) and the evidence against models with $\sigma = 0.02$ was very weak (*N2:* $BF_{01} = 1.3$, *spill-over:* $BF_{01} = 1.3$).

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. We further found evidence for a slowdown at N2 and the spill-over region in NP2 attachment sentences. As in experiment 1, this finding likely reflects late RC attachment in the NP2 attachment condition.

Importantly, our data did not indicate the presence of an effect of a magnitude which was comparable to previously attested instances of an ambiguity advantage. We had hypothesized that such an effect would occur either at N2 or at the spillover position. The posterior probability of an ambiguity advantage of the smallest expected magnitude at N2 was very low (0.005), and relatively low (0.09) at the spill-over region. Moreover, the Bayes Factor analysis provided a moderate level of evidence against nearly all models allocating a substantial amount of probability mass to an effect size between 3% and 19%. The only models for which the Bayes Factors in favor of the absence of an effect were *anecdotal* (but positive) $\sigma = 0.04$ and $\sigma = 0.02$, which allocated a relatively large amount of probability mass to effect sizes below 3% (.77 and .93, respectively), and a relatively low amount of probability mass to effect sizes above 3% (.23 and .07, respectively). These findings suggest that either the structure studied in this experiment does not elicit an ambiguity advantage or if it does, its effect size is most likely substantially smaller than previously attested instances.

General Discussion

We conducted two reading experiments to test a prediction made by the strategic underspecification account: underspecification of attachment in structures with prenominal relative clauses. If readers strategically underspecify ambiguous sentences, they should do so even when RCs are prenominal, as they are in Turkish. Therefore, under the assumption of strategic underspecification, we expected to observe an ambiguity advantage in prenominal relative clauses. Under a race-based account such as the SMCM, however, no ambiguity advantage was expected.

In Experiment 1 (N=36), 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-cut 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 NP1 attachment conditions, the regression path duration as well as the number of first-pass regressions pointed to a potential, albeit small speedup. A Bayes Factor analysis failed to provide conclusive evidence for or against a difference in reading times between ambiguous and N1 attachment conditions, although it did provide tentative evidence against a *substantial* speedup. 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 with a different experiment design and a substantially increased number of participants (N=184). We found no evidence for the presence of an ambiguity advantage. Moreover, additional analyses based on (i) the posterior of the estimate of the difference between the ambiguous and N1 attachment conditions, and (ii) on a Bayes Factor analysis provide evidence against the presence of an ambiguity advantage of the magnitude that has previously been reported for English ($22\,ms - 120\,ms$).

These findings present a challenge for both interpretations of the underspecification account by Swets et al. (2008). When task demands do not encourage disambiguation, *early underspecification* predicts an ambiguity advantage at NP2. This is because RC attachment is initially underspecified and later resolved in the NP2 attachment condition, but remains underspecified in the NP1 attachment condition. This should lead to a speedup in the ambiguous condition relative to the unambiguous NP1 condition. *Late underspecification* predicts that the parser delays the process of attaching the RC until such a point where it knows with certainty whether the structure is ambiguous. Based on that information, it decides to underspecify RC attachment in ambiguous conditions, and to attach the RC in unambiguous attachment conditions. This is expected to happen either at the second noun or at the spill-over region. Since our findings in Experiment 2 reflect either the absence of an ambiguity advantage or a significantly reduced effect size, they pose a challenge for the underspecification account. If they reflect the absence of an effect, the findings are at odds as they indicate the absence of underspecification in structures with prenominal RCs. The presence of an ambiguity advantage of a reduced magnitude, on the other hand, has two possible interpretations: Either underspecification occurs, but significantly less often than in English, thus leading to a decreased average effect, or the relative speed-up due to the underspecification of RC attachment is significantly smaller in Turkish than in English. In either case, is not currently clear what factors are responsible for the the difference between postnominal and prenominal structures.

Importantly, race-based models such as the SMCM (Logačev & Vasishth, 2016) do not predict an ambiguity advantage in ambiguous structures with prenominal RCs because they assume that RC attachment begins at the earliest opportunity, which means that an attempt to attach the RC to NP2 must begin with a significant delay relative to the attempt to attach it to NP1. As a result, the chance that attachment to NP2 will be completed first is negligible, and no ambiguity advantage is expected. Thus, our findings are compatible with the SMCM, but pose a number of challenges for the underspecification account.

Conclusion

We proposed two different interpretations of the Swets et al. (2008) strategic underspecification account and discussed how the ambiguity advantage follows from their specific sets of assumptions. We argued that in contrast to race-based theories, both implementations of the strategic underspecification account predict an ambiguity advantage not only in languages with postnominal RCs, like English, but also in languages with prenominal RCs, such as Turkish. We further presented the results of two reading time experiments which failed to show an ambiguity advantage comparable in magnitude with previously attested instances of the ambiguity advantage. We argue that our findings present a major challenge for the strategic underspecification account, but are compatible with race-based models such as the SMCM.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390–412.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi: 10.18637/jss.v080.i01

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

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

Dinçtopal-Deniz, N. (2010). Relative clause attachment preferences of turkish l2 speakers of english. *Research in second language processing and parsing*, 27–63.

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. doi: 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. doi: 10.3115/1073336.1073357

Jeffreys, H. (1961). *Theory of probability (3rd edn.)*. Oxford, UK: Oxford University Press.

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. doi: 10.1037/a0020693

Kırkıcı, B. (2004). The processing of relative clause attachment ambiguities in turkish. *Turkic Languages*, *8*(1), 111–121.

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

Lee, M., & Wagenmakers, E. (2014). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge: Cambridge University Press.

Lenth, R. V. (2021). emmeans: Estimated marginal means, aka least-squares means [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=emmeans` (R package version 1.7.0)

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

Nicenboim, B., & Vasishth, D. S. S. (2021). *An introduction to bayesian data analysis for cognitive science*. Retrieved from `https://vasishth.github.io/bayescogsci/book/index.html`

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Raab, D. H. (1962). Statistical Facilitation of Simple Reaction Times. *Transactions of the New York Academy of Sciences*, *24*(5 Series II), 574–590. doi: 10.1111/j.2164-0947.1962.tb01433.x

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic bulletin & review*, *25*(1), 128–142.

Schotter, E. R., Bicknell, K., Howard, I., Levy, R., & Rayner, K. (2014). Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition*, *131*(1), 1–27.

Stan Development Team. (2020). *RStan: the R interface to Stan.* (R package version 2.21.2)

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*(1), 201–216. doi: 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. doi: 10.1006/jmla.1998.2600

van Gompel, R. P., 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. doi: 10.1016/j.jml.2004.11.003

van Gompel, R. P., 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. doi: 10.1016/B978-008043642-5/50029-2

van Gompel, R. P., 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. doi: 10.1006/jmla.2001.2773

Vasishth, S. (2021, Aug). *A common mistake in data analysis (in psychology/linguistics): Subsetting data to carry out nested analyses (part 1 of 2).* Retrieved from `https://vasishth-statistics.blogspot.com/2021/08/a-common-mistake-in-psychology-and.html`

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147–161.

von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of memory and language*, *94*, 119–133.

von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1545–1578. doi: 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. doi: 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. doi: 10.1080/17470218.2017.1307862

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Wickham, H. (2017). tidyverse: Easily install and load the 'tidyverse' [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=tidyverse` (R package version 1.2.1)

Wotschack, C., & Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *The Quarterly Journal of Experimental Psychology*, 15.

Zehr, J., & Schwarz, F. (2018). Penncontroller for internet based experiments (ibex) [Computer software manual]. Retrieved from `https://doi.org/10.17605/OSF.IO/MD832`