A multinomial processing tree model of RC attachment

Pavel Logačev
Boğaziçi University
Department of Linguistics
34342 Istanbul, Turkey
pavel.logacev@gmail.com

Noyan Dokudan
Boğaziçi University
Department of Linguistics
34342 Istanbul, Turkey
noyan.dokudan@gmail.com

Abstract

Studies in the field of sentence processing often use a variant of a discrete choice task to determine speakers' understanding of ambiguous sentences. We discuss a largely neglected source of potential bias and variability that may affect the results of such experiments. We propose an analysis method based on multinomial processing tree models (Batchelder and Riefer, 1999) which corrects for this bias and allows for a separation of parameters of theoretical importance from nuisance parameters. We test two variants of the MPT-based model. We further demonstrate that this method can provide deeper insight into the processes underlying participants' answering behavior as well as their attachment preferences.

1 Introduction

One of the key questions in the field of sentence processing has been: *What does the human sentence processing mechanism do when confronted with an ambiguity?* A variety of different proposals for a variety of disambiguation strategies have been made over the years, such as the Garden-path Theory (Frazier, 1987), the Tuning Hypothesis (Cuetos et al., 1996), the Competition-Integration Model (McRae et al., 1998) and many others. Their diverging predictions have led to a large body of empirical research documenting, among other things, substantial cross-linguistic variation in the interpretation of ambiguous sentences: For instance, Cuetos and Mitchell (1988) compared the RC attachment preferences of English and Spanish speakers in ambiguous sentences like (1) and (2), in which the relative clause *'who had the accident'* can attach either to the NP headed by the first noun (N1, *'daughter'*) or to the NP headed by the second noun (N2, *'colonel'*).[1]

Cuetos and Mitchell presented Spanish-speaking and English-speaking participants with ambiguous sentences like (1) and (2) and asked them to answer comprehension questions like *'Who had an accident?'*. Participants' responses indicated that English sentences like (1) received an N2 interpretation in 61% of the cases, while their Spanish counterparts like (2) received an N1 interpretation in 72% of the cases. The authors interpret this finding as an argument for a universal parsing strategy in the resolution of RC attachment ambiguities.

(1) The journalist interviewed the daughter$_{N1}$ of the colonel$_{N2}$ [who had an accident].

(2) El periodista entrevisto a la hija$_{N1}$
The journalist interviewed to the daughter
del coronel$_{N2}$ [que tuvo el accidente].
of the colonel [who had an accident].

Although the speakers' disambiguation strategy seems to be at least in part determined by language-specific factors, a variety of other factors appear to influence ambiguity resolution in RC attachment. In an off-line questionnaire study, Gilboy et al. (1995, inter alia) demonstrated a substantial influence of construction type. They asked participants to indicate which of the two available noun phrases was modified by the RC in several constructions. They found that the percentage of N2 attachment responses ranged between approximately 20% to 70% for their English sentences, and between 10% to 80% for their Spanish sentences. Grillo et al. (2015) also conducted a two-alternative forced choice (2AFC) task in which English speakers choose between N1 and N2 as the attachment sites for the RC to indicate their interpretation of sentence. They showed that English speakers, who had previously been claimed to prefer N2 attachment, preferred N1 attachment in more than 50% of the cases when a small clause reading was possible.

[1] In order to avoid any ambiguity in the context of typologically diverse languages, we will refer to the two interpretation options as *N1 attachment* and *N2 attachment*, with N1 and N2 referring to the order of occurrence of the noun phrases head nouns instead of the more common terms *high attachment* and *low attachment*.

RC attachment preferences have also been studied in Turkish, where the order of RC and complex noun phrase is reversed. In a questionnaire study with sentences like (3), Kırkıcı (2004) found that animacy may affect attachment preferences such that when both NPs are [+human], there are no significant difference between the proportions of the N1 and N2 attachment, while an N1 attachment manifested when both NPs were [-human]. Contrary to this finding, Dinçtopal-Deniz (2010) found an across-the-board preference for N1 attachment in Turkish. In her questionnaire study, monolingual Turkish speakers read Turkish sentences with ambiguous RC attachment and answered questions about those by indicating one of two options on each trial. The results of this study showed that participants preferred N1 attachment over N2 attachment (66% and 34%, respectively).

(3) Şoför [şehir merkezinde
driver in the city center living
oturan]$_{RC}$ profesörün$_{N1}$ sekreterini$_{N2}$
professor's secretary saw
gördü.

'The driver saw the secretary of the professor who was living in the city center.'

2 The Role of Guessing

What most of the above studies of RC attachment preferences have in common is that they use some variant of a discrete choice task, in which participants select one of two response options to indicate their interpretation of the ambiguity. The relative proportion of responses indicating N1 and N2 attachment, respectively are interpreted as estimates of the magnitude of N1 or N2 attachment. A potential complication in interpreting the percentages of N1 and N2 responses in this way is that participants' responses may not always reflect their interpretation. It appears quite likely that, on some trials, participants process the sentence only partially or fail to pay attention to it altogether. In such cases, participants' question responses must be based on an incomplete or nonexistent representations, and are more likely to resemble guesses than informed responses.

Evidence for such incomplete processing comes from the widely known fact that participants' accuracy in experimental tasks is often far from perfect, even for relatively simple tasks such as acceptability judgments: For example, Dillon and Wagers (2019) found in an *off-line* acceptability judgment study that ungrammatical sentences like (4) are judged acceptable on 18% of the trials. Since it appears unlikely that sentences like 4 are considered grammatical and interpretable when fully processed, the explanation for such responses must lie in their incomplete processing followed by guessing.

(4) *Who do you think that the new professor is going to persuade anyone?

One way of conceptualizing a simple generative model of erroneous responses in relatively simple tasks is to assume that at least some participants on some occasions fail to pay attention to the stimulus, and as a result, select a random response.

If so, the relation between the probability of response *X* being actually preferred to alternative responses (p_X) and the probability of observing response *X* (p'_X) can be formalized as in equation 1: p'_X is the weighted average of (i) the probability of *X* being preferred to the alternative when the stimulus is fully processed (p_X) and (ii) the probability of selecting *X* when the stimulus is not attended (g_X), where a is the probability of attending to the stimulus.

$$p'_X = a \cdot p_X + (1 - a) \cdot g_X \qquad (1)$$

Given that participants in most if not all psycholinguistic tasks produce a sizeable amount of erroneous responses, it appears *a priori* quite plausible that such mechanisms are also at play in attachment preference studies. This means that empirical estimates of attachment preferences (p'_X) are likely to be (i) *biased* towards the guessing parameter g_X to a degree determined by a, and (ii) are likely to *vary between studies* as a function of both, a and g_X.

In the following, we propose a method for disentangling the contributions of attachment preferences and guessing using multinomial processing tree models (MPT; Erdfelder et al., 2009; Batchelder and Riefer, 1999) based on response patterns in unambiguous baseline sentences. We present two alternative MPT models for the task of answering polar comprehension questions about sentences with ambiguous and unambiguous RC attachment in two languages, English and Turkish.

3 Experiments

To evaluate our method, we used question-answering data from two experiments in which participants read sentences with ambiguous and unambiguous RC attachments and answered polar comprehension questions about them.

3.1 Experiment 1

We used the question-answering data from the RC questions condition in Swets et al.'s (2008) self-paced reading experiment in English (N=48), in which participants read sentences like (5) in three attachment conditions and answered comprehension questions about RC attachment which were similar to (6) on every trial. All comprehension questions required a *'yes'/'no'* answer. One-half of the questions asked whether the RC modified the noun phrase headed by N1, and the other half asked about N2.

RC attachment was disambiguated by means of gender (mis)match between the reflexive in the RC and the RC head noun. Each participant read 36 experimental sentences. Unambiguous sentences had correct answers, while the responses to ambiguous sentences indicate how readers disambiguated the sentence, thus reflecting their RC attachment preference.

Figure 1 (left panel) shows the average percentages of *'yes'* responses to comprehension questions by attachment condition and question type (question about N1 or N2).

(5) a. AMBIGUOUS ATTACHMENT
The maid$_{N1}$ of the princess$_{N2}$ [who scratched *herself* in public] ...
b. N1 ATTACHMENT
The son$_{N1}$ of the princess$_{N2}$ [who scratched *himself* in public] ...
c. N2 ATTACHMENT
The son$_{N1}$ of the princess$_{N2}$ [who scratched *herself* in public] ...
... was terribly humiliated.

(6) COMPREHENSION QUESTION
Did the maid/princess/son scratch in public?

3.2 Experiment 2

The second set of question-answering data came from an unpublished self-paced reading experimental about RC attachment in Turkish (N=99). In an experiment design similar to Swets et al., participants read sentences like (7). Because Turkish relative clauses are pre-nominal, the RC *who hit each other* preceded the complex noun phrase *the fans of the football players*. RC attachment of the was disambiguated by the number marking of the head nouns as RCs with the reciprocal *each other* can only modify plural noun phrases. The structure is ambiguous if both nouns are plural because they are both licit attachment sites for the RC. Participants were asked *'yes'/'no'* comprehension questions about RC attachment, like (8). The comprehension question asked about the event mentioned in the RC and whether one of the nouns was involved in that event. Each participant read 42 experimental sentences. One-half of the questions asked whether the RC modified the noun phrase headed by N1, and the other half asked about N2. The experiment was conducted online on *ibexfarm* (Drummond, 2013). All participants were undergraduate students at Boğaziçi University and were native speakers of Turkish.

Figure 1 (right panel) shows the average percentages of *'yes'* responses to comprehension questions by attachment condition and question type (question about N1 or N2).

(7) Dün akşam, [birbirini döven]$_{RC}$ ...
Yesterday evening, each other hit
a. AMBIGUOUS ATTACHMENT
futbolcu-lar-ın$_{N1}$ hayran-lar-ı$_{N2}$ ...
footballer-PL-GEN fan-PL-POSS
b. N1 ATTACHMENT
futbolcu-lar-ın hayran-ı ...
footballer-PL-GEN fan.SG-POSS
c. N2 ATTACHMENT
futbolcu-nun hayran-lar-ı ...
footballer.SG-GEN fan-PL-POSS
... stadyumu hemen terk etti.
stadium immediately leave did.
'The fans of the football players who hit each other left the stadium immediately, yesterday evening.'

(8) COMPREHENSION QUESTION
Futbolcu(lar)/hayran(lar) dövüşte yer almış mı?
'Was/were the football player(s)/fan(s) involved in the fight?'

The average percentages of *'yes'* responses in figure 1 indicate a substantial number of errors in unambiguous experimental conditions, such as *'no'*-responses to N1 questions and *'yes'*-responses to N2 questions about N1 attachment sentences.

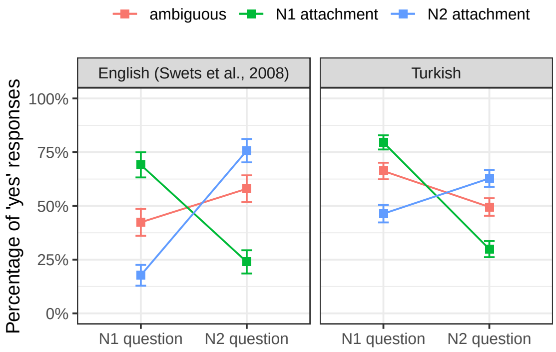

Figure 1: Average percentates of *'yes'* responses by attachment condition (color) and question type (x-axis). Error bars indicate 95% within-subject CIs.

The average accuracy in answering questions about unambiguous sentences was 79% ($SE = 1.3$) in Swets et al.'s English experiment, and 66.5% ($SE = 2.5$) in the Turkish experiment. The difference is likely due to the fact that the former was conducted in a lab, while the latter was conducted online.

The question responses in ambiguous attachment conditions indicate an N2 attachment preference in the English as 58% ($SE = 2.1$) of the response are compatible with N2 attachment (*'yes'* responses to N2 questions and *'no'* responses to N1 questions). Meanwhile, the Turkish data indicates an N1 preference as 58% ($SE = 1.9$) of the question responses are compatible with an N1 interpretation of the sentence. In both cases, the preferred attachment option is *local*, i.e., adjacent to the RC, which is consistent with prior research.

In spite of the fact that the estimates of the magnitude of the attachment preference are coincidentally equal, these numbers cannot be taken at face value because of the presence of a substantial number of erroneous responses in unambiguous conditions in both experiments. These responses suggest that not all N1- or N2-compatible responses indicate that the participant has successfully formed an N1- or N2 attachment interpretation of the sentence, as they may have been generated by the same extraneous cognitive process that generates erroneous responses in unambiguous attachment conditions. In the next section, we test two models of erroneous responses and then use them to quantify the actual strength of the attachment preference.

4 MPT models of question-answering and attachment

In accounting for the influence of extraneous cognitive processes, we considered two mechanisms that may generate erroneous question responses, and implemented both as multinomial processing tree (MPT) models (Batchelder and Riefer, 1999) in order to the model with the better empirical fit to obtain less biased estimates of the attachment preferences in the ambiguous conditions, as indicated by question responses.

MPT models offer a way to formalize hypotheses about how a mixture of several latent processes generates a categorical response (cf. Erdfelder et al., 2009, for an overview). Events involved in processing are represented as a probability tree, with hypothesized latent processes either occurring or not occurring at each junction, with each path corresponding to unique combinations of cognitive processes which give rise to a particular response, along with the probabilities of each path. Importantly, this formalization provides a framework in which the probabilities of relevant latent processes can be estimated. We will use them to estimate the magnitude of the RC attachment preference in Turkish and English.

4.1 Model 1

The first mechanism we considered as a potential explanation for erroneous responses is that people simply fail to attend to the stimulus or fail to perform the task of answering questions and simply press a random button. The logic of this account is illustrated in figure 2. The MPT schematic on top illustrates how events during processing of an N1 attachment sentence can unfold: On any given N1 attachment trial, a participant may be in an attentive state (probability: a), or in an inattentive state (probability $1 - a$).

If the participant is in an attentive state throughout the trial (i.e., during reading and question answering), they will form a memory trace of the sentence they read, and later use it to correctly answer comprehension questions. This is illustrated in the top branch of the N1 attachment condition MPT schematic in figure 2, where *'N1 response'* stands for *'yes'* responses to N1 questions and *'no'* responses to N2 questions.

If the participant is in an inattentive state at some point during the trial (i.e., during reading or question answering), they will either fail to form a memory trace of the sentence they read or will fail to use it to answer the comprehension question. On those occasions, they will respond *'yes'* with probability g, and *'no'* with probability $1 - g$. This is illustrated in the bottom branch of the N1 attachment condition MPT schematic in figure 2.

As a result of these assumptions, the probability of a *'yes'* response in the N1 attachment condition is as given in equation 2, where I_{N1} is an indicator variable which is 1 for N1 comprehension questions and 0 for N2 comprehension questions.

The processing assumptions for the N2 attachment (middle, figure 2) condition and the ambiguous condition (bottom, figure 2) follow a similar logic, with the probability of a *'yes'* response given by equations 3 and 4. One important assumption about the hypothesized process for question-answering in ambiguous attachment conditions is that when readers are in an attentive state, they disambiguate ambiguous sentences either towards an N1 interpretation (with probability h) or an N2 interpretation (with probability $1 - h$). We make no assumptions about whether that happens during reading or at the question answering stage.

$$p_{Y|N1} = a \cdot I_{N1} + (1 - a) \cdot g \tag{2}$$

$$p_{Y|N2} = a \cdot (1 - I_{N1}) + (1 - a) \cdot g \tag{3}$$

$$p_{Y|A} = a \cdot [h \cdot I_{N1} + (1 - h) \cdot (1 - I_{N1})] + (1 - a) \cdot g \tag{4}$$

Importantly, we make no claims as to what may bring on inattentiveness and when exactly it occurs (i.e., during reading or questions answering). We do speculate, however, that it may be brought on by distractions in the environment, mind-wandering ((e.g., Smallwood, 2011)), or (temporary) fatigue. However, an important prediction of this account is that this process should affect the different attachment conditions to the same degree, as its rate shouldn't depend on the attachment condition.

4.2 Model 2

A second possible source of erroneous responses may be processes that do not affect all conditions equally across the board. For example, it could be that one of the two interpretations is more often successfully created during reading, or more often successfully recalled during question answering. Moreover, given that RC attachment is disambiguated via anaphor binding in both experiments,

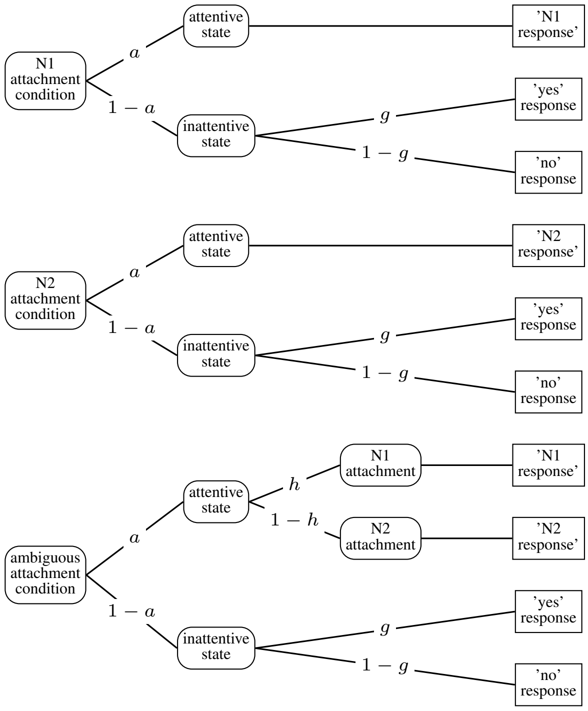

Figure 2: An MPT model of question answering with equal error rates for (i) N1 attachment, (ii) N2 attachment, and (iii) ambiguous sentences.

it is, in principle, possible that the preferred interpretation is usually attempted first, even in dispreferred attachment conditions, before that interpretation is rejected, and the correct dispreferred interpretation is constructed. As a result, the presence of a partial alternative structure could interfere with the retrieval of the correct structure during question answering (e.g., Staub, 2007).

We formalized the assumption of different error rates associated with N1 and N2 attachment in the MPT model in figure 3. The hypothesized structure of unambiguous N1 and N2 attachment trials is similar to model 1 in figure 2. Each attachment process is associated with a probability of complete recollection certainty, r_1 and r_2, which reflect the probability that the correct sentence structure is (i) constructed during reading and (ii) later correctly recalled during the question-answering phase. If the correct sentence structure is constructed and later recalled with certainty, participants respond in accordance with the structure they constructed. Otherwise, they select a random response, i.e., *'yes'* with a probability of g and *'no'* with a probability of $1 - g$. The probability of a *'yes'* response for all attachment conditions is given in equations 5, 6, 7.

Importantly, in the ambiguous condition (figure

3, bottom), the recollection certainty and recollection uncertainty nodes are nested under the RC attachment nodes because the probabilities of the recollection certainty and uncertainty states depend on which RC attachment was chosen. One way to interpret this model structure is that while there are (i) causes of recollection uncertainty which apply to all conditions equally and which affect processing across the board with probability $a \leq min(r_1, r_2)$, as well as (ii) additional attachment condition-specific causes of recollection uncertainty which affect processing with probability $r_1 - a$ in when an N1 attachment is adopted, and with probability $r_2 - a$ when N2 attachment adopted. However, all these sources of recollection uncertainty are assumed to lead to the same guessing behavior. Thus, the model in 3 subsumes the model in 2.

$$p_{Y|N1} = r_1 \cdot I_{N1} + (1 - r_1) \cdot g \quad (5)$$

$$p_{Y|N2} = r_2 \cdot (1 - I_{N1}) + (1 - r_2) \cdot g \quad (6)$$

$$p_{Y|A} = h \cdot p_{Y|N1} + (1 - h) \cdot p_{Y|N2} \quad (7)$$

5 Method

We implemented both MPT models in *brms* and *rstan* (Bürkner, 2018; Stan Development Team, 2020) in R (R Core Team, 2018) according to equations 2-7. We fitted the models to each experiment separately, using 4 MCMC chains with $1,000$ warm-up and $3,000$ post-warm-up iterations. For the sake of computational convenience, we estimated all model parameters on the logit scale, and in the following, we will use θ' to refer to the logit-transform of a parameter θ.

We used mildly informative Gaussian priors for all logit-transformed parameters in both models: $h', g' \sim N(0, 1)$, and $a, r_1, r_2 \sim N(0, 1)$.

To account for individual differences in all parameters, we used hierarchical models with by-subject intercepts for all parameters, where each participant k's responses were modeled as a function of population-level parameters θ with subject subject-level adjustments $\delta_{\theta,k}$, with $\theta'_k = \theta' + \delta_{\theta',k}$, where the by-subject adjustments are distributed as $\delta_{\theta',k} \sim N(0, \sigma_{\theta'})$.

6 Model Comparison

Figures 4 and 5 show the average percentages of *'yes'* responses by experiment alongside 95% pos-

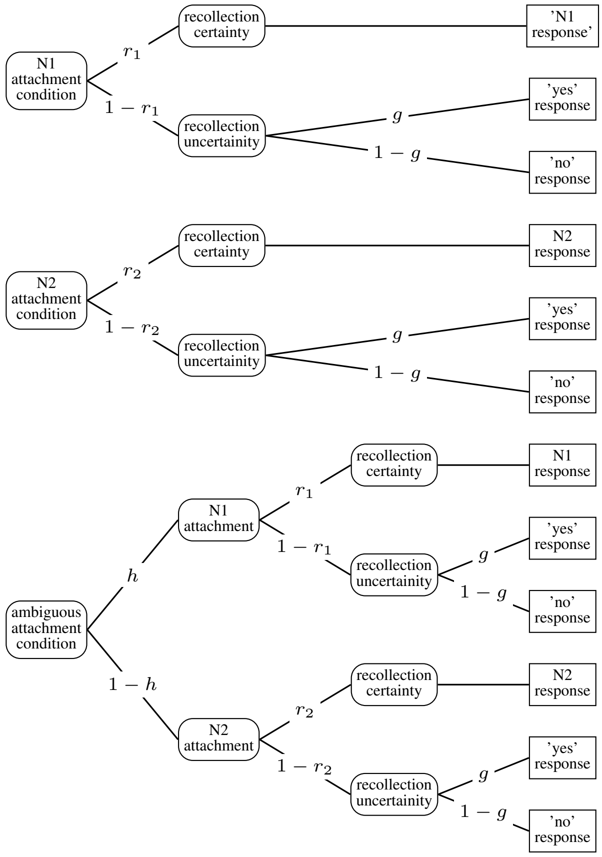

Figure 3: An MPT model of question answering with different error rates for N1 attachment and N2 attachment processes.

terior predictive intervals generated by, models 1 and 2, respectively.

Figure 4 shows that although model 1 could approximate the experimental findings to a degree, it systematically overestimated the proportion of responses compatible with the preferred RC attachment (N2 in English, N1 in Turkish) in both unambiguous conditions: For example, in the N1 attachment condition in English, the number of *'yes'* responses to N1 questions and *'no'* responses to N2 questions were slightly overestimated. Similarly, in the N2 attachment condition in Turkish, the model underestimated the percentage of *'yes'* responses to N1 questions and *'no'* responses to N2 questions. Figure 5 shows that model 2 appears to have fewer systematic deviations, and appears to fit the data quite well.

In order to compare the candidate models more formally, we computed each model's expected log pointwise predictive density ($\widehat{elpd}$) using PSIS-LOO-CV (Vehtari et al., 2017) as an estimate of the model's out-of-sample performance. Table 1 shows the $\widehat{elpd}$ estimates, the differences between

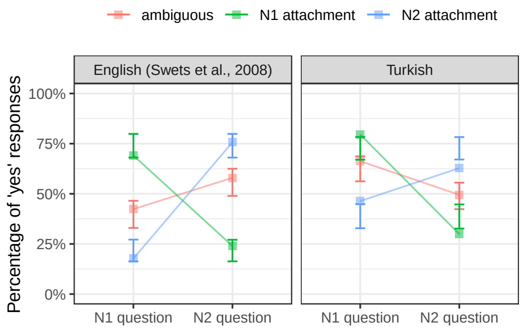

Figure 4: Average percentages of *'yes'* responses and 95% posterior prediction intervals based on model 1 by attachment condition and question type.

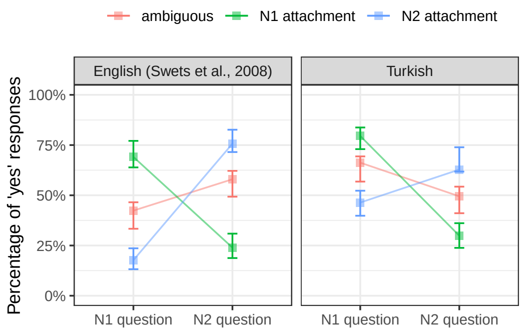

Figure 5: Average percentages of *'yes'* responses and 95% posterior prediction intervals based on model 2 by attachment condition and question type.

	English	Turkish
	$\widehat{elpd}$	$\widehat{elpd}$
model 1	-511.3 (18.1)	-796.4 (15.0)
model 2	-469.4 (14.8)	-750.5 (13.5)
	$\Delta\widehat{elpd}$	$\Delta\widehat{elpd}$
model 2-1	41.9 (11.6)	45.9 (9.3)

Table 1: Estimates of expected log pointwise predictive density ($\widehat{elpd}$) by model for each experiment and differences between model $\widehat{elpd}$s. Standard errors in brackets.

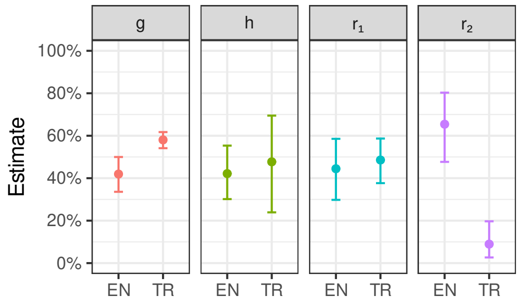

Figure 6: Population parameter estimates and 95% credible intervals for all four parameters of model 2 (g, h, r_1, r_2) for both experiments, English (EN) and Turkish (TR).

models in $\widehat{elpd}$ as well as their respective standard errors. Given that both $\Delta\widehat{elpd}$ estimates are relatively large relative to their standard errors, they both point towards model 2 as having better out-of-sample performance. This finding suggests that the two attachment processes are affected by the error-generating process to different degrees.

7 Discussion

Having established model 2 as an adequate model of RC attachment in the context of question-answering, we used its parameter estimates to understand the pattern of responses than accuracy rates: Figure 6 shows the population parameter estimates and 95% credible intervals for all four parameters of model 2. In addition to difference in the guessing bias g between experiments, it also shows a rather small difference between experiments in the estimates of attachment preference parameter h, which represents the probability with which the parser adopts an N1 attachment interpretation over an N2 attachment structure in ambiguous attachment conditions. The estimates of the N1 attachment probability were 42% ($CrI = [30; 55]$) for the English experiment and 48% ($CrI = [24; 70]$) for the Turkish experiment. While the estimate for English is consistent with an N2 attachment preference, the estimate for Turkish indicates no clear preference.

But why does the estimate of the attachment preference parameter h for Turkish lead to different conclusions than the percentage of N1 responses in the ambiguous condition? Generally speaking, the discrepancy is due to the fact that the percentage of N1 response conflates several parameters. In this particular instance, the explanation seems to lie in the very low estimate of the parameter r_2 (9%, $CrI = [3; 20]$), which indicates that N2 attachment structures were unlikely to be successfully constructed or recalled. A possible reason for this is that even on trials resulting in an N2 interpretation, the parser always attempts to construct an N1 attachment structure first because, in Turkish, unlike in English, potential attachment sites are processed sequentially. As a result, a remnant of the N1 attachment interpretation may interfere with the retrieval of the memory trace of the N2 attachment interpretation. It appears likely, that as a result of retrieval failure, participants would resort to guessing, which in turn should lead to a substantial number of erroneous responses following N2 attachment sentences or ambiguous sentences which were disambiguated towards N2 attachment.

Whatever really the true source of higher error rates in the N2 attachment conditions in the Turkish experiment, our MPT analysis suggests that what appears as a weak N1 attachment preference in our Turkish experiment is actually a consequence of a large number of guessing trials associated with N2 attachment. It appears that when an N1 attachment structure is adopted in ambiguous sentences, participants often give an N1 response on approximately 50% of the trials and guess on the remaining trials. In contrast, when an N2 attachment structure is adopted, participants tend to correctly give an N2 response on only approximately 10% of the trials and guess on the remainder. Assuming that N1 structures are adopted approximately 50% percent of the time, the expected proportion of N1 responses in the ambiguous condition is 60%, which is close to the 58% we obtained.

In sum, our analysis shows that (i) the N2 attachment preference in the English experiment appears to hold up even when guessing trials are taken into account, and (ii) that what appears to be an N1 attachment preference is readily explained by processing difficulty associated with processing and recalling N2 attachment structures in Turkish.

8 Summary

In this paper, we have discussed a previously neglected source of bias and variability that may affect studies of attachment preferences and interpretation preferences more generally. We proposed an MPT-based analysis method which allows to de-confound parameters of theoretical importance from nuisance parameters such as the guessing rate. We tested two variants of the MPT-based model, and demonstrated that this method can provide further insight into the processes underlying participants' answering behavior as well as their attachment preferences.

References

William H. Batchelder and David M. Riefer. 1999. Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, 6(1):57–86.

Paul-Christian Bürkner. 2018. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1):395–411.

Fernando Cuetos and Don C. Mitchell. 1988. Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, 30(1):73–105.

Fernando Cuetos, Don C. Mitchell, and Martin Corley. 1996. Parsing in different languages. In *Language Processing in Spanish*. Erlbaum, Hillsdale, NJ.

Brian Dillon and Matthew Wagers. 2019. Approaching gradience in acceptability with the tools of signal detection theory. Preprint, Open Science Framework.

Nazik Dinçtopal-Deniz. 2010. Relative clause attachment preferences of Turkish L2 speakers of English: Shallow parsing in the L2? In Bill VanPatten and Jill Jegerski, editors, *Language Acquisition and Language Disorders*, volume 53, pages 27–64. John Benjamins Publishing Company, Amsterdam.

A. Drummond. 2013. Ibex farm.

E. Erdfelder, T. Auer, B. Hilbig, A. Aßfalg, M Moshagen, and L Nadarevic. 2009. Multinomial Processing Tree Models: A Review of the Literature. *Zeitschrift für Psychologie / Journal of Psychology*, 217(3):108–124.

Lyn Frazier. 1987. Sentence Processing: A Tutorial Review. In Max Coltheart, editor, *Attention and Performance XII*. Erlbaum, Hillsdale, NJ.

Elizabeth Gilboy, Josep-MMaria Sopena, Charles Cliftrn, and Lyn Frazier. 1995. Argument structure and association preferences in Spanish and English complex NPs. *Cognition*, 54(2):131–167.

Nino Grillo, João Costa, Bruno Fernandes, and Andrea Santi. 2015. Highs and Lows in English Attachment. *Cognition*, 144:116–122.

Bilal Kırkıcı. 2004. The processing of relative clause attachment ambiguities in Turkish. *Turkic Languages*, 8:111–121.

Ken McRae, Michael J. Spivey-Knowlton, and Michael K. Tanenhaus. 1998. Modeling the Influence of Thematic Fit (and Other Constraints) in Online Sentence Comprehension. *Journal of Memory and Language*, 38(3):283–312.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

J Smallwood. 2011. Mind-wandering While Reading: Attentional Decoupling, Mindless Reading and the Cascade Model of Inattention: Mind-wandering While Reading. *Language and Linguistics Compass*, 5(2):63–77.

Stan Development Team. 2020. RStan: the R interface to Stan. R package version 2.19.3.

A. Staub. 2007. The return of the repressed: Abandoned parses facilitate syntactic reanalysis73. *Journal of Memory and Language*, 57(2):299–323.

Benjamin Swets, Timothy Desmet, Charles Clifton, and Fernanda Ferreira. 2008. Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, 36(1):201–216.

A. Vehtari, Gelman A., and Gabry J. 2017. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27.