

The Role of Shallow Processing in Agreement Attraction

Utku Türk, Pavel Logačev

Boğaziçi University

Abstract

We report the results of two speeded acceptability judgment experiments in Turkish. We hypothesized an alternative explanation for agreement attraction effects in Turkish that is based on shallow processing. Our findings contradict our hypothesized form-driven processing strategy and support an account of agreement attraction based on the use of abstract linguistic features, rather than mere form.

The Role of Shallow Processing in Agreement Attraction

sectionIntroduction ¹.

Agreement Attraction Thingy

Morphophonology.

related sub-sub-field studies.

This work details.

Experiment 1

In Experiment 1, we tested whether speakers of Turkish make fewer errors when the local ambiguity in the head noun is resolved. Specifically, we have replicated the Lago et al. (2018) study. However, instead of using their NP heads, we chose to use nouns that end in a vowel, thus resolving the ambiguity between the accusative and possessive.

Participants

One hundred seven native speakers of Turkish (X male, Y female) in Experiment 1. The mean age is Z , ranging from $Z1$ to $Z2$. Experiments in this paper were carried out following the Declaration of Helsinki and the regulations concerning ethics in research in Boğaziçi University. All participants provided informed consent. They were tested online, on IbexFarm, using their preferred platform,

Materials

In Experiment 1, participants first saw sentences word by word in the middle of their screen. As in Lago et al. (2018) study, we used 40 sets of item and manipulated the grammaticality of the sentences (*grammatical* x *ungrammatical*) and the number of the attractor noun (*singular* x *plural*). All experimental sentences followed the following template: *Noun*_{GEN}-*Noun*_{POSS}-*Adverb*-*Verb*. The distribution of the verb types followed

¹

the same pattern as in Lago et al. (2018), namely twenty unergatives, eighteen unaccusatives, and two optionally transitive verbs. The matrix verb was marked with overt plural marker *-lar* in the half of experimental items. Adverbials in our experimental sentences were always pre-verbal and contained 2-3 words consisting of 15 characters on average. The second noun phrase is the head noun of the subject in all of our experimental sentences whereas the first noun phrase is the genitive modifier of the head, thus *the attractor*. Unlike Lago et al. (2018), our head nouns were vowel-ending, which enables us to distinguish between accusative (*-yI*) and possessive (*-sI*). This change was due to the morpho-phonologic nature of possessive suffix in Turkish, which spell-outs as (*-I*) after a consonant same as the accusative. While all of the head nouns were singular, the attractor noun was in its plural form in the half of the experimental items. One example set of experimental items can be seen in (1).

(1) a. *Plural Attractor, Ungrammatical Verb*

*Yönetici-ler-in aşçı-sı mutfak-ta sürekli zıpl-ıyor-lar.
 manager-PL-GEN cook-POSS kitchen-LOC non-stop jump-PROG-PL.
 ‘The cooks of the manager were jumping in the kitchen non-stop.’

b. *Singular Attractor, Ungrammatical Verb*

*Yönetici-nin aşçı-sı mutfak-ta sürekli zıpl-ıyor-lar.
 manager-GEN cook-POSS kitchen-LOC non-stop jump-PROG-PL.
 ‘The cook of the manager were jumping in the kitchen non-stop.’

c. *Plural Attractor, Grammatical Verb*

Yönetici-ler-in aşçı-sı mutfak-ta sürekli zıpl-ıyor.
 manager-PL-GEN cook-POSS kitchen-LOC non-stop jump-PROG.
 ‘The cooks of the manager was jumping in the kitchen non-stop.’

d. *Singular Attractor, Grammatical Verb*

Yönetici-nin aşçı-sı mutfak-ta sürekli zıpl-ıyor.
 manager-GEN cook-POSS kitchen-LOC non-stop jump-PROG.
 ‘The cook of the manager was jumping in the kitchen non-stop.’

As for the filler sentences, we have included forty sentences, half of which ends in

singular and ungrammatical while the other half was plural and grammatical. These decisions were made due to the non-existence of such sentences in the experimental sentences. We wanted to rule out a possible task strategy where participants decide on the grammaticality of the sentence by merely checking the final plural marking. Even though filler items did not follow a strict template as the experimental items, they were created within a template of *Noun_{GEN}-Noun_{POSS}-Converb-Noun-Adverb-Verb*. The ungrammaticality was created non-locally in the filler items to avoid task strategies. Example filler sentences can be seen in (2).

(2) a. *Plural Grammatical Verb*

Sosyolog-un öğrenci-si konuş-unca tutarsızlık açığ-a
sociolog-GEN student-POSS speak-NMLZ inconsistency open-DAT
çıkardılar.
deduct-PST-PL.

‘When the sociologist’s student talked, they exposed an inconsistency.’

b. *Singular Ungrammatical Verb*

*Dansöz-ün koca-sı var-ınca kapı sakince aç-tı.
dancer-GEN husband-POSS arrive-NMLZ door slowly open-PST.

Intended: ‘When the husband of the dancer came, the door opened slowly.’

Procedure

Prior to the experiments, participants gave information regarding themselves and consent to participate in the experiment. They read the instructions where we showed what they will see in the experiment, and where we informed them to give quick and accurate answers during the experiment. They are also informed regarding the that there will be a time-limit in answering the question. Every session started with a practice session where participants have been informed to be a bit faster when they were not. The sentences were presented in the middle of the screen word-by-word on a web-based platform Ibex Farm (<http://spellout.net/ibexfarm/>). Each word was displayed with XX font size and stayed on the screen for 300 ms, unlike Lago et al. (2018) and Wagers, Lau, and Phillips (2009) where words stayed for 500 ms in the

screen. After the sentence is completed, participants were asked: "How did they find the sentence?" They were informed to press P on their keyboard for *good* and Q for *bad*. Apart from the practice items, we did not provide any feedback regarding their answers. However, participants were instructed to be faster when they waited more than 5,000 ms to answer the acceptability question.

Participants saw forty experimental and forty filler items. Experimental items were distributed among four different lists using Latin Square design. Every participant saw one version of the experiment with a specific list and one item per condition.

Analysis

Every experimental item from both experiments (Experiment 1 and Lago et al. (2018)) is included in the analysis. Participants ($n=X$) that exceeded the threshold of 0.25 difference in giving accurate answers to grammatical conditions (1-d) and (1-b) are excluded from the analysis. We used R packages *brms* (Bürkner, 2018) and *rstan* (Stan Development Team, 2019) to fit Bayesian hierarchical models (Gelman & Hill, 2007). We added our independent variables, i.e. grammaticality of the sentence and the overt number information on the distractors, and random variable experiment as a factor. Analyses by items and participants are also conducted as a part of the Bayesian hierarchical models. We added our independent variables, i.e. grammaticality of the sentence and the overt number information on the distractors, and random variable experiment as a factor. Analyses by items and participants are also conducted as a part of the Bayesian hierarchical models. Data regarding our experiments and analysis can be found in `link hidden due to blind review`.

Results

Figure 2 shows the average proportions of ‘acceptable’ responses by experimental condition for both the original experiment in Lago et al. (2018) and our replication with unambiguous possessive marking. It shows that ungrammatical sentences with plural attractors are rated as acceptable more often than their counterparts with singular attractors. In addition to ungrammatical conditions, accuracy rates within the

grammatical conditions were comparable and high across conditions (mean = x%, SE = x%). As seen in Figure 2, participants judged the sentences with plural attractor and singular verbs to be grammatical with the rate of X%, which is in line with the findings reported in Lago et al. (2018).

The estimates and 95% CIs of a Bayesian GLM in Figure 3 confirm the observations we stated above: the positive interaction between sentence grammaticality and attractor number is in line with a larger effect of attractor number in ungrammatical sentences. The main effect of ungrammaticality was XX in log-odds, and the interaction between the ungrammaticality and plural attractor was XX in log-odds. Since other estimates' credible intervals range over zero, we conclude that they were not informative.

Discussion

In Experiment 1, we have found number attraction effects in Turkish using genitive possessive structures, as it is also attested in Lago et al. (2018). To account for their findings, Lago et al. (2018) assume a cue-based memory retrieval mechanism. That is, they assume that upon reaching the verb, the parser attempts to retrieve its agreement controller (the subject) using a cue-based retrieval mechanism (Jäger, Engelmann, & Vasishth, 2017; Lewis & Vasishth, 2005). The assumption is that in sentences such as (1), features such as case and number information are used to identify the agreement controller in memory. In ungrammatical sentences, when the verb bears plural agreement, no NP in memory will match both retrieval cues. However, in ungrammatical plural attractor conditions, the attractor matches one of the cues, which can lead to its erroneous retrieval on some occasions. This cannot happen in ungrammatical singular attractor conditions. This difference in the probability of erroneous retrievals is presumably what surfaced as a number agreement attraction effect, as observed in Lago et al. (2018), and our Experiment 1.

In their work, Lago et al. (2018) argue that the agreement attraction in genitive-possessive structures in Turkish is due to the use of genitive case as a marker of

embedded subjects in Turkish, i.e. differential subject marking (DSM) properties of Turkish (Kornfilt, 2009). They argue that the genitive case in Turkish may not provide strong cue against subjecthood due to DSM while it does in English. Following this logic, we hypothesized that the other phenomenon that gives a strong clue against the subject should inhibit the illusionary dependencies as argued in Nicol, Barss, and Barker (2016). One such phenomenon was present in the experimental items used in Lago et al. (2018): the morpho-phonologic ambiguity between the accusative and the possessive case. Since they use only consonant ending heads, the marking on the head is ambiguous between the accusative and the possessive marking. Following their argumentation, participants may search for other agreement controllers than the possessive marked heads when they are engaged in shallow processing and erroneously assign head as genitive marked distractor.

Nevertheless, we found that when the possessive marker is disambiguated agreement attraction does not diminish in effect size. As seen in Figure 3, we successfully replicated the findings of Lago et al. (2018) with disambiguated head nouns. Thus, we conclude that the morpho-phonological possessive-accusative ambiguity plays no role in number attraction in Turkish.

However, there is an alternative explanation that has yet to be ruled out: task-specific strategies. We hypothesized that readers may engage in an even shallower process in the evaluation of the sentences such as (3) in which readers decide on the acceptability of the sentence with a faulty state of memory. In our model, we claim that after they read the sentence readers may end up with insufficient information to reliably judge the sentence. Upon such cases, readers may end up using extremely shallow processing methods such as matching the agreement-wise unrelated but form-wise related elements. We present our hypothesized decision tree in Figure 4

The aim of our second experiment was to test whether agreement attraction in Turkish may be an instance of a *form-driven processing strategy*. Assuming that readers sometimes engage in shallow processing, they may end up with insufficient information to reliably classify a sentence as (un)acceptable. In such cases, participants may choose

to classify sentences with plural-agreement-bearing verbs as acceptable if they have a memory of a nominal plural morpheme in the sentence. Such a response strategy would lead to a larger number of ‘acceptable’ responses in ungrammatical plural attractor conditions than in ungrammatical singular attractor conditions.

Experiment 2

In Experiment 2, we wanted to rule out the possibility of *form-driven processing strategy*, namely participants deciding on the acceptability of a sentence using a memory of plural morpheme in the sentence when they do not have sufficient information to rate sentences (un)acceptable.

Participants

Eighty native speakers of Turkish (X male, Y female) in Experiment 1. The mean age is Z , ranging from $Z1$ to $Z2$. Experiments in this paper were carried out following the Declaration of Helsinki and the regulations concerning ethics in research in Boğaziçi University. All participants provided informed consent. They were tested online, on IbexFarm, using their preferred platform,

Materials

In Experiment 2, participants saw the sentences word by word in the middle of their screen. We have again formed 40 sets of items. The grammaticality of the sentences (*grammatical* x *ungrammatical*) and the number of the attractor (*singular* x *plural*) were manipulated. Unlike Experiment 1, we used nominalized relative clause attractors instead of nouns. We took advantage of syncretism between Turkish nominal and verbal plural marker. Both of these morphemes spells-out as *-lar*, which enables us to check whether agreement attraction in Turkish can be explained by an extremely shallow dependency parsing based on the forms of morphemes rather than linguistic features.

All experimental sentences followed the same template as the experiment one except for the nature of the attractor: *RC-Bare Noun-Adverb-Verb*. We have used the

same verbs, and have not changed the distribution of verb types. We also utilized the same or extremely similar adverbials in length. We again did not manipulate the number of the head noun and manipulated the number of the attractor. Relative clauses we used in this experiment are all object relative clauses and they are all marked with canonical *-dIK* nominalizer. Since Turkish is a pro-drop language, we also dropped the subject within the embedded clause, thus ending up with a one-word object relative clause whose head is also the controller of the number agreement on the matrix verb. One example set of experimental items can be seen in (3).

(3) a. *Plural Attractor, Ungrammatical Verb*

*Tut-tuk-lar-ı aşçı mutfak-ta sürekli zıpl-ıyor-lar.
 hire-NMLZ-PL-POSS cook kitchen-LOC non-stop jump-PROG-PL.
 ‘The cook that they hired were jumping in the kitchen non-stop.’

b. *Singular Attractor, Ungrammatical Verb*

*Tut-tuğ-u aşçı mutfak-ta sürekli zıpl-ıyor-lar.
 hire-NMLZ-POSS cook kitchen-LOC non-stop jump-PROG-PL.
 ‘The cook that they hired were jumping in the kitchen non-stop.’

c. *Plural Attractor, Grammatical Verb*

Tut-tuk-lar-ı aşçı mutfak-ta sürekli zıpl-ıyor.
 hire-NMLZ-PL-POSS cook kitchen-LOC non-stop jump-PROG.
 ‘The cook that they hired was jumping in the kitchen non-stop.’

d. *Singular Attractor, Grammatical Verb*

Tut-tuğ-u aşçı mutfak-ta sürekli zıpl-ıyor.
 hire-NMLZ-POSS cook kitchen-LOC non-stop jump-PROG.
 ‘The cook that they hired was jumping in the kitchen non-stop.’

In our filler items for Experiment 2, we made sure that every sentence starts with an object relative clause. We used plural distractors with grammatical verbs and singular distractors with ungrammatical verbs. In all of our filler sentences, the dependency between the first NP subject and its verb resolved in an adverbial embedded sentence. Grammatical filler items in Experiment 2 all had a template of RC_{PL} -Bare

Noun-Adverb-Converb-Noun-Adverb-Verb_{PL}, whereas ungrammatical filler items used a template of *RC_{SG}-Bare Noun-Adverb-Converb-Noun-Adverb-Verb_{SG}*

Again, to avoid a possible strategy where participants use plural ending as a direct indication of ungrammaticality, half of our fillers were with an overt plural marking on a grammatical verb while the other half were without an overt plural marking on an ungrammatical verb. We used Turkish pro-drop characteristics which enable participants to form a dependency between the matrix verb and the null subject. Example filler sentences can be seen in (4)

(4) a. *Plural Grammatical Verb*

Oku-t-tuk-lar-ı öğrenci başarılı ol-unca mutlu ol-du-lar.
read-CAUS-NMLZ-PL-POSS student successful be-NMLZ happy be-PST-PL.

‘When the student they sponsored become successful, they became happy.’

b. *Singular Ungrammatical Verb*

*Kandır-dığ-ı adam öde-me-yince bulaşık saatlerce yıka-dı.
trick-NMLZ-POSS man pay-NEG-NMLZ dish for.hours clean-PST.

Intended: ‘When the man he tricked did not pay, he cleaned dishes for hours.’

Procedure and Analysis

In Experiment 2, the same procedure as Experiment 1 has been followed with regards to how sentences are presented within the experiment.

We analyzed the percentages of ‘yes’ answers given by the participants in experimental conditions. In the analysis of this experiment, we have included experimental items from our Experiments 1 and 2 to check the interaction between experiments and quantify the difference between verbal and nominal plural markers. In Experiment 2 as well, we excluded participants ($n=X$) who had more than 0.25 points of difference in accuracy between grammatical conditions, which implicates that they did not give enough attention to all conditions. We used R packages brms (Bürkner, 2018) and rstan (Stan Development Team, 2019) to fit Bayesian hierarchical models (Gelman & Hill, 2007). We added our independent variables, i.e. grammaticality of the

sentence and the overt number information on the distractors, and random variable experiment as a factor. Analyses by items and participants are also conducted as a part of the Bayesian hierarchical models. Data regarding our experiments and analysis can be found in `link hidden due to blind review`.

Results

Figure 5 shows the average proportions of 'acceptable' responses by experimental conditions for both Experiment 1 and Experiment 2. We see that there is no effect resembling agreement attraction in the ungrammatical conditions of Experiment 2. Ungrammatical conditions showed comparable effects (X, Y) regardless of the attractor's number.

The observations we made through Figure 5 also surfaces when we look at the estimates and 95% CIs of a Bayesian GLM in Figure 6. As is visible from the figure, the model shows a negative three-way interaction between grammaticality, type of attractor (nominal vs. verbal), and attractor number, which entails a reduced effect of agreement attraction in Experiment 2 compared to Experiment 1.

Discussion

In Experiment 2, we could not find any effect of agreement attraction in the acceptability ratings of ungrammatical conditions with a verbal plural attractor. However, number agreement attractions were observed when the attractors were nominal. Even though the forms of the two plural morphemes are the same form-wise, there is a distinct difference between the results of two experiments. One of the indications of zero-effect with verbal plural morpheme is that readers do not make decisions solely based on the form of the elements. This finding contradicts our hypothesized form-driven processing strategy and supports an account of agreement attraction based on the use of abstract linguistic features, rather than mere form.

However, results may be more telling with regards to the other theories of agreement attraction. These results seem unexpected within the cue-based retrieval mechanisms. Even though the root of the attractor is verbal, the morpheme is the same

with the third person plural possessive marker since the relative clauses in Turkish are always nominalized and the subject is marked with the genitive case. The findings of this experiment indicate towards more fine-grained features needs to be utilized. One other possible explanation within the cue-based retrieval models can be formed around the status of agreement markers. One can say that the agreement process is completed in the verbal *-lar*; in other words, it is not the triggering morpheme but the probe of the agreement. This difference would mean that verbal *-lar* should not start any search for dependency resolution. In contrast to verbal *-lar*, the plural morpheme on the genitive is supposed to be triggering part of the agreement, not the probe. However, in Turkish embedded clauses, where genitive subjects can be seen, the number agreement works differently than the matrix level agreement. When the genitive subject of an embedded clause is marked with plural it is ungrammatical to have plural marking on the verb. This prohibition in return means that the genitive subjects should not create an expectation for a plural marked verb, which was believed to be the reason for the agreement attraction in Lago et al. (2018).

As for the marking and morphing theories, the results of Experiment 2 are completely expected and mitigated by the syntactic depth effects. Due to the genitive possessor's limited inner complexity compared to that of relative clause modifiers, marking and morphing theories would expect higher contribution from the genitive attractors to the final representation of number information.

General Discussion

References

- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1), 395–411. doi: 10.32614/RJ-2018-017
- Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press.
- Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, 94, 316–339. doi: 10.1016/j.jml.2017.01.004
- Kornfilt, J. (2009). Dom and two types of dsm in turkish. In *Differential subject marking* (pp. 79–111). Springer.
- Lago, S., Felser, C., Gračanin-Yukse, M., Demir, O., Kırkıcı, B., & Şafak, D. F. (2018). Straight from the horse’s mouth: Agreement attraction effects with Turkish possessors. *Linguistic Approaches to Bilingualism*, 12(1), 1–29. doi: 10.1075/lab.17019.lag
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419. doi: 10.1207/s15516709cog0000_25
- Nicol, J. L., Barss, A., & Barker, J. E. (2016). Minimal interference from possessor phrases in the production of subject-verb agreement. *Frontiers in psychology*, 7, 548.
- Stan Development Team. (2019). *RStan: the R interface to Stan*. Retrieved from <http://mc-stan.org/> (R package version 2.19.1)
- Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2), 206–237. doi: 10.1016/j.jml.2009.04.002

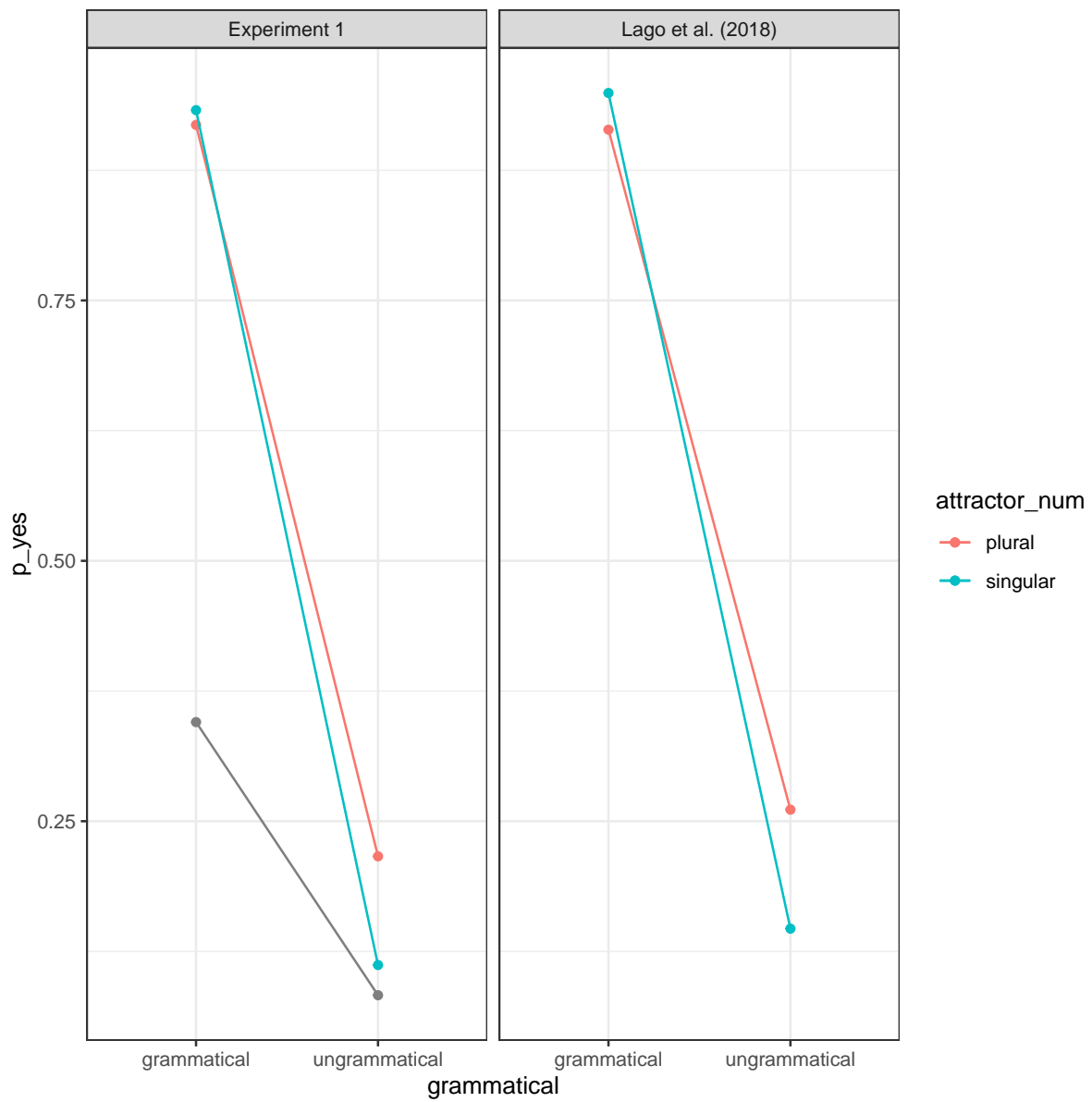


Figure 1. Scatter Plot with Regression Line

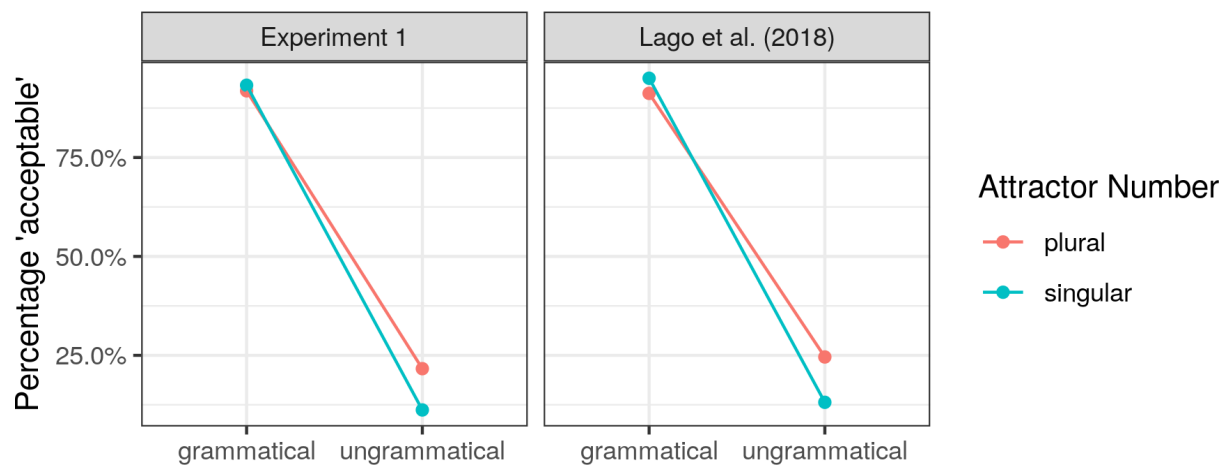


Figure 2. Percentage of 'yes' (acceptable) responses of Experiment 1 and Lago et al. (2018)

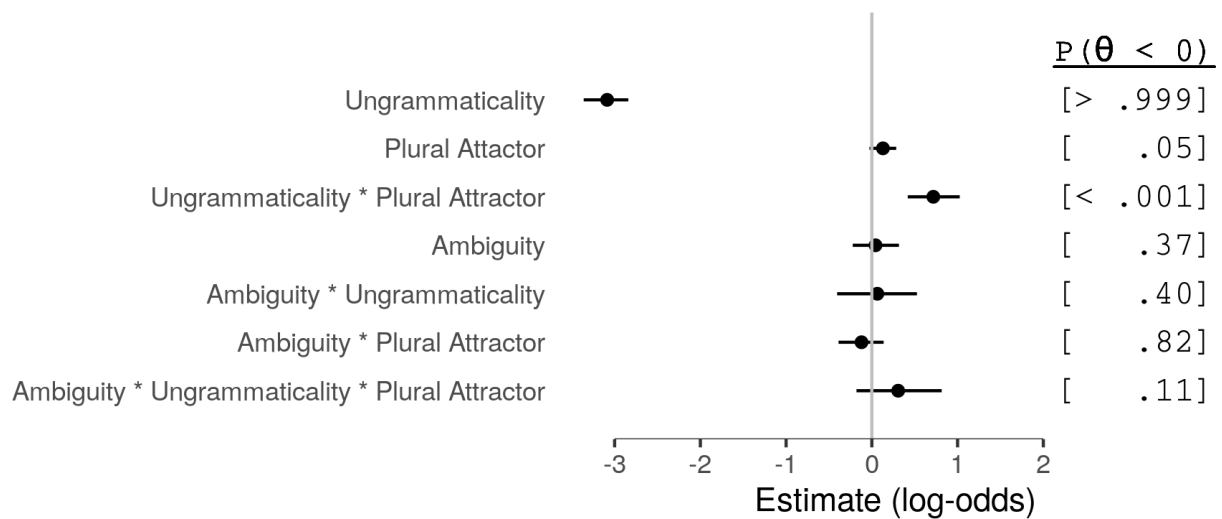


Figure 3. Estimates and 95% credible intervals for the regression coefficients for Experiment 1

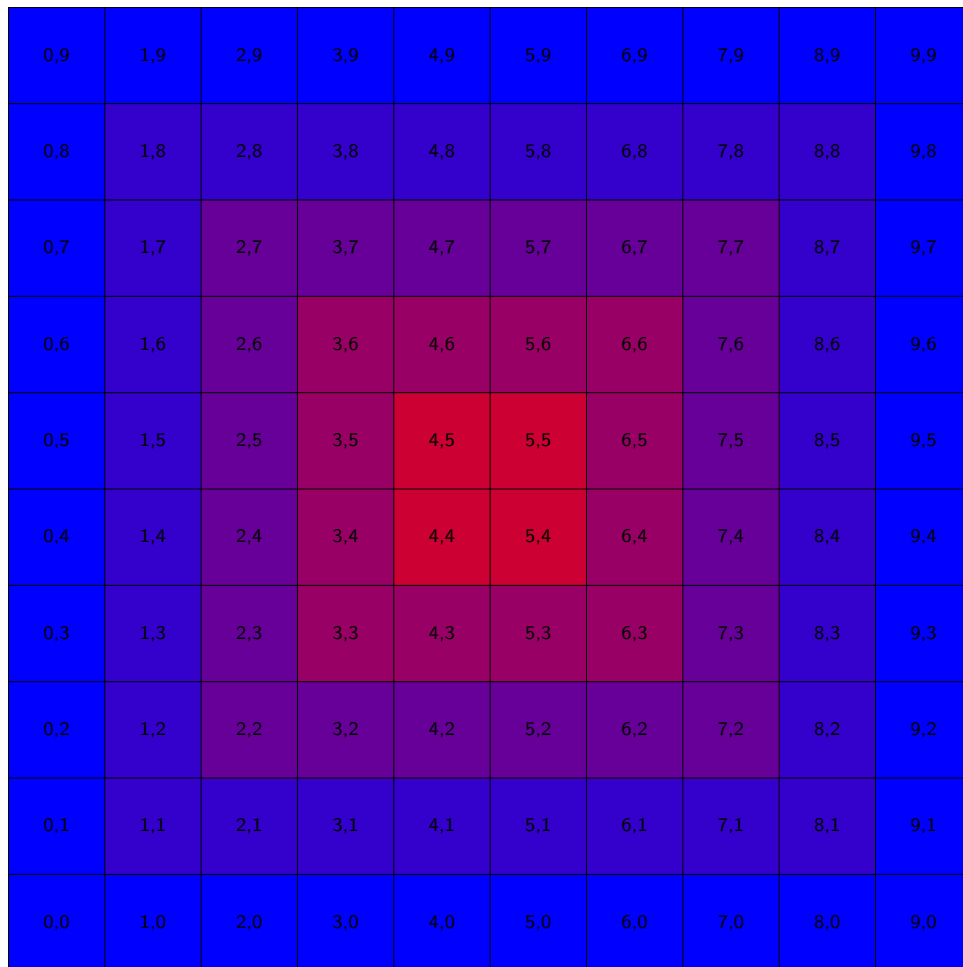


Figure 4. Proposed multinomial processing tree of how people judge sentences in an agreement attraction task

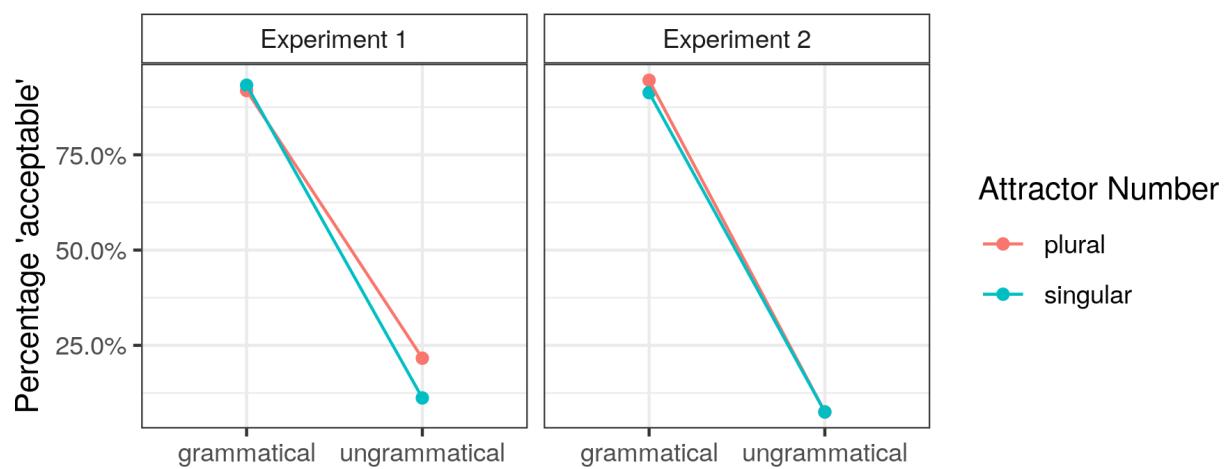


Figure 5. Estimates and 95% credible intervals for the regression coefficients for Experiment 1 and Experiment 2

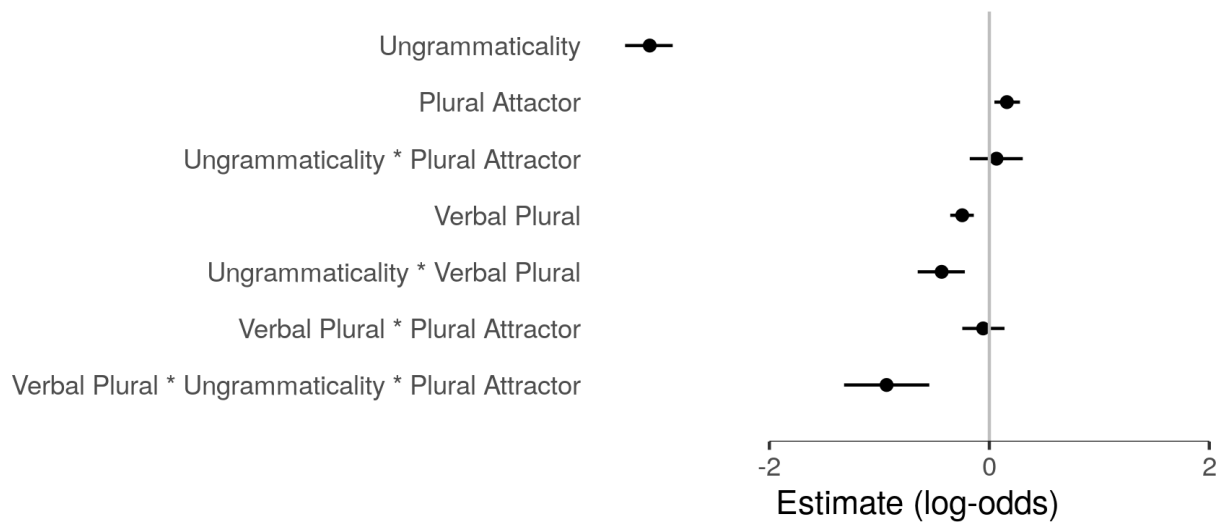


Figure 6. Estimates and 95% credible intervals for the regression coefficients for Experiment 2