

Authors' Response to Reviews of Agreement Attraction in Turkish: The Case of Genitive Attractors

Utku Türk and Pavel Logačev
LCN

RC: Reviewers' Comment, AR: Authors' Response, □ Manuscript Text

Dear Editor and Reviewers,

We thank you for your hardwork in preparing your reviews and giving us the opportunity to submit a revised version of our manuscript entitled *Agreement Attraction in Turkish: The Case of Genitive Attractors*. We highly appreciate the constructive criticisms and valuable comments of the anonymous reviewers on our manuscript. We have revised the manuscript in the submitted version based on the comments of the reviewers.

Please find below our response to the reviewers' comments. We hope the reviewers find the revised manuscript and our response below to satisfactorily address their comments.

Please note that reviewers' comments are in bold and italics while our answers are not. Additions to the original manuscript are indicated in blue and deletions with red *only in the response letter*.

For reference, we add our previous submission file to the end of this response letter in which line numbers are shown. We also include an additional PDF in which all additions are indicated with red font.

Yours Sincerely,
Utku Türk, Department of Linguistics, Boaziçi University
Pavel Logačev, Department of Linguistics, Boaziçi University

1. Reviewer #1

1.1. Major Points

RC: *The findings in the article could be more broadly related to other work on the role of morphological syncretism in agreement attraction. As it stands, the paper focuses too narrowly on Lago et al. (2019). Even though the confound in Lago et al.s study is about case-marking syncretism, there is no discussion on the role of syncretism on agreement attraction. As a result, the final conclusion comes across as a bit flippant: "Participants do not appear to rely on form-related cues in their decision-making processes and instead use of abstract linguistic features as retrieval cues."*

This conclusion would be justifiable if the current article were the only one that had looked at the role of form-related features (i.e., syncretism) on agreement attraction. But this issue has been examined before, and there seems to be evidence that case syncretism modulates attraction, both in production (Hartsuiker, Schriefers, Bock, and Kikstra, 2003; Hartsuiker, Antón-Méndez, and van Zee, 2001) and in comprehension (Slioussar, 2018; Cherepovskaia, Reutova, & Slioussar, 2021). Writing "form-related cues dont

modulate attraction” ignores all this work.

I know that this is a brief article, so I am not suggesting that the introduction is expanded to describe all these previous studies. But acknowledging them would strengthen the introduction and would also make the discussion more nuanced. For example, the authors could mention in the introduction that their suspicion that syncretism may have played a role in Lago et al. (2019) is consistent with previous work showing that case ambiguity modulates attraction rates (with references to previous work). And in the discussion, they might want to avoid generic statements of the type “form-related cues don’t modulate attraction” and instead speculate on why they did not modulate attraction rates in their study (see my suggestion below regarding the possible infrequency of the alternative structural option).

todo

todo

1.2. Minor Points

1.2.1 Page 1, line 44

RC: *The keywords “Bayesian GLM” and “replication” may be either too general or misleading. The manuscript uses Bayesian GLMs as a tool. Bayesian models are not the focus of the article and no specific inferences or conclusions are specifically based on the method (the effects are really clear, so I suspect a frequentist analysis would have revealed similar patterns). In that context, using “Bayesian GLM” as a keyword may be misleading for readers (with that logic, most psycholinguistic articles should list “frequentist mixed-models” as a keyword). Meanwhile, the article is not a replication of Lago et al. (2019) in a strict sense, since it changed the materials (if the study had found different patterns, we would not call it a “failed replication”, I think). So perhaps the two keywords could be replaced for more discipline-specific terms, e.g., “sentence processing” or “comprehension”, “number”, etc.*

AR: Following the reviewer’s comments, we have changed the keywords.

Agreement attraction; syncretism; Turkish; sentence processing; number ~~Bayesian GLM~~; ~~replication~~

1.2.2 Page 2, lines 44-45

RC: *“in the experimental sentences in their experiment” > “in their experimental sentences”.*

AR: Following the reviewer’s comments, we have changed the wording.

... which is related to an instance of local ambiguity in their ~~the~~ experimental sentences ~~in their~~ experiment ...

1.2.3 Page 3, lines 34-35

RC: *“in which the verb agreed” > “in which the verb agreed in number” (agreement could be in principle with any number of features, e.g., person).*

AR: Following the reviewer’s comment, we specified what the verb agrees with.

... in which the verb agreed in number with the singular subject head noun ...

1.2.4 Page 3, lines 53-55

RC: “from case and number information is” > are? (Perhaps this is an instance of agreement attraction?)

AR: It was indeed an instance of agreement attraction. Following the reviewer’s comments, we corrected the agreement.

... how case and number information are ~~is~~ encoded and retrieved ...

1.2.5 Page 4: Quantitative Acc vs Poss

RC: *The motivation for the experiment is that a sequence such as “Teknisyen-in egitmen-i” is ambiguous, such that the consonant-ending NPs second NP could be interpreted as the head of a possessive phrase (the structural parse assumed by Lago et al., 2019) or as the object of an embedded clause. I am wondering whether it’s possible to get some quantitative estimate of this ambiguity.*

Specifically, would it be possible to estimate, given such sequences in a corpus, how often they resolve to each type of structure? Such an estimation would allow inferring whether case syncretism plays no role (the interpretation favored by the authors) or whether it did not have an observable effect in the study because the alternative parse the second NP as the object of an embedded clause is a very infrequent one.

We thank the reviewer for their comments on quantifying the probabilities of possible parses. We now report the relative probability of having an accusative marked noun following the genitive-marking. In calculating our estimate, we have used 5 treebank from the Universal Dependencies Project: UD Turkish Kenet, UD Turkish IMST, UD Turkish GB, UD Turkish FrameNet, UD Turkish BOUN, all of which add up to 38,237 sentences.

In our analysis, we divided the count of accusative marking following an instance of genitive marking by total count of genitive marking preceded by accusative and possessive marking.

$$P(Acc|Gen) = \frac{C(Acc|Gen)}{C(Acc|Gen) + C(Poss|Gen)}$$

Our changes can be found right before example (4).

We also calculated the relative probability of having an accusative marked noun compared to an possessive marked noun following genitive marking. Data from annotated treebanks from Universal Dependencies v2.9 (Cöltekin, 2015; Kuzgun et al., 2020; Sulubacak et al., 2016; Türk et al., 2021, 2019) showed that the relative probability of encountering accusative marking after the genitive-marked noun is 0.21.

1.2.6 Page 5, line 6

RC: *example (4b): translation error? What is “the fast”? Shouldn’t it be “the instructor”?*

AR: Following the reviewer’s comment, we corrected the translation error in the (4b).

I saw the technician firing the instructor fast .

1.2.7 Page 5, line 40

RC: *“are completely anonymised” > was?*

AR: Following the reviewer’s comment, we corrected the agreement and tense.

All participants provided informed consent before their participation and their identity was ~~are~~ completely anonymised.

1.2.8 Page 7: Save Space

RC: *if you need to save space in order to expand the theoretical discussions, the description of the fillers on page 7 (“We hypothesized that....”) could be shortened. The concern raised is pretty standard in psycholinguistics, so it’s perhaps not necessary to devote a whole paragraph to explaining it. Instead, you could briefly say something like “To prevent participants from equating singular verbs with grammaticality, 40 fillers were added featuring 20 with grammatical plural verbs and 20 with ungrammatical singular verbs”.*

todo

todo

1.2.9 Page 9, line 10

RC: *“experiment compared” > “experiment and compared”*

Following the reviewer’s comment, we included the conjunction.

... we analyzed the data from the present experiment and compared our results to Lago et al.’s (2019) results ...

1.2.10 Page 9: Data Exclusion

RC: *Analysis section: excluding more than 10% of the experimental trials seems a lot to me: typically, in my own studies on attraction (as well in the other papers I review) researchers exclude 5% of their data or less.*

I also find it potentially problematic to exclude participants based on the performance in the experimental trials themselves. If the experimental manipulation plays a role, it may affect precisely response properties like the difference in “yes” responses in grammatical vs. ungrammatical conditions (the first criterion listed in the text).

I think that it is more justifiable to use the fillers to assess a “yes” response bias. This is one of the purposes of having fillers, and since 1/2 the fillers were grammatical, and 1/2 were not, this procedure is feasible in the current study. Thus, my suggestion is to only exclude participants based on their filler performance. If the authors decide against this (or if the number of excluded trials exceeds 5% even with a filler-based exclusion criterion), then it would be useful to report a supplementary analysis showing that the results don’t qualitatively change when the data with all participants is taken into account. This doesn’t have to be in the main text, it could a supplementary material, or part of the response letter.

We agree with the reviewer that 10% exclusion is indeed more than what has been the case in other works. We now exclude only 3.39% of the participants. We did so by lowering the accuracy threshold from 25% to 10% and eliminating the RT threshold.

We believe that there should be an accuracy threshold based on the difference in “yes” responses in grammatical vs. ungrammatical conditions **with singular attractors**. We believe that we should have specified that we are using only the singular attractor conditions more clearly. In our original manuscript, we only refer to those items. Now, we specifically included “**with singular attractors**” phrase.

Specifically, we removed all participants for whom the difference in the percentage of yes responses between conditions with singular attractors (5c) and (5d) ~~the grammatical condition (5d) and the ungrammatical condition (5c)~~ fell below the threshold of ~~0.25~~ 0.10 percentage points. We also excluded trials in which the participants missed the response deadline ~~or gave overly fast responses (below 200 ms)~~. As a result, we excluded ~~10.22%~~ 3.39% of the trials from our experiment and 2.38% of the Lago et al.s (2019) trials.

1.2.11 Page 9: Priors

RC: *“We have used standard priors provided by the brms package” > This statement is not very informative. And as defaults may change across time, it also makes the analysis less reproducible. I would suggest reporting the actual priors used.*

As the reviewer pointed out, standard brms priors indeed change over time. Instead of the standard priors, we now specified "semi-informative" priors. We specified these priors explicitly. For the changes, reviewers may check page 3rd paragraph of the analysis section. We also provide the last version below.

~~We have used standard priors provided by the brms package.~~ A Student’s t -distribution ($\nu = 3$, $\mu = 0$, $\sigma = 2.5$), a normal distribution ($\mu = 0$, $\sigma = 1$), a Cauchy distribution ($(\mu = 0, \sigma = 1)$) and a LKJ distribution ($\eta = 2$) were used for our intercept, slopes of predictors, standard deviations of random effects, correlation coefficients in interaction models respectively as priors.

1.2.12 Page 9, lines 39-41

RC: *“We used by-participant and by-item intercepts and slopes for all predictors.” And their interaction?*

Following the reviewer's comment, we included the term "and their interaction" as it was already present in the analysis itself.

We used by-participant and by-item intercepts and slopes for all predictors [and their interaction](#) .

1.2.13 Page 10: Error Bars

RC: *"Error bars signal standard errors calculated following Cousineau et al. (2005); Morey et al. (2008)." > Rather than citing 2 different papers, it would be more useful to just briefly state how they were calculated, especially if they are simply SEs across participants and items. Is it realistic to expect readers to go look up two different papers in order to interpret error bars?*

todo

todo

2. Reviewer #2

2.1. Engagement with case and agreement attraction literature

RC: *Let me preface this by saying: I understand that the 3,000 word format provides a strong limit on how much discussion of anything can be undertaken. However, the two claims in the final paragraph of the paper deserve a bit more context. Here they are repeated for reference:*

"[...] our results suggest (i) that agreement attraction effects in Turkish are not due to a reduced association between case-ambiguous nouns and the subjecthood features, and (ii) that local ambiguities, such as case syncretism, do not appear to play a role in agreement attraction."

For example, a number of studies have found that if the attractor noun bears nominative case, or a case that is syncretic with nominative, there is increased attraction errors. This has been shown in Slovak (Badecker and Kuminiak, 2007), Russian (Slioussar, 2018), German (Hartsuiker et al., 2003), and Dutch (Hartsuiker et al., 2001). More relevant to the current study, Nicol and Antón-Méndez (2009) showed that in English, replacing the attractor with the unambiguous accusative pronoun them lead to fewer attraction errors than the full DP, which is ambiguous between nominative and accusative. This is very much akin to the present manipulation, but with "opposite" findings. The authors should ensure that all of these papers are cited and discussed to provide the context for the present result. I believe only Badecker and Kuminiak (2007) are currently cited, and this is just in passing with no reference to their findings regarding case. I recommend that the authors think carefully about how their results fit into these previous work. Likely, this will require a weakening of the statements above. Though, the authors may find the results of Avetisyan et al (2020), which also should be discussed in more detail, to dovetail with the results of the current study. In short, this complicated picture about the relationship between case, case syncretism, and agreement attraction should be made more clear.

todo

todo

2.2. The notion of subject

RC: *I think the authors should carefully consider each instance of the use of the term “subject”. I would highly recommend reading the paper McCloskey (1997) “Subjecthood and Subject Positions”, which provides a clear and critical overview of the notion of “subject”. This is a bigger issue than just this paper I believe it has been a huge mistake within this literature to appeal to a feature [Subject], when the nature of subjecthood turns out to be impossible to define in any singular way. In other words, we call things that have a certain set of properties “the subject”, but this can differ quite greatly from language-to-language or construction-to-construction. I believe the authors have an opportunity to get away from this type of feature and be more precise about what they mean by “subject”. For example, on page 4 when the authors mention “genitive NPs can function as subjects of embedded clauses”, in what sense do they mean this? In terms of being an agreement controller? In terms of the syntactic position it occupies? Of course, it is not in the sense that these NPs bear nominative case (they are genitive). I think this matters a lot, because cue-based retrieval accounts are only as good as the features they are operating over. Given the dubiousness of [Subject] as a grammatical primitive, every effort should be made to be clear about what is actually meant when we use the term (i.e. the controller of agreement, the noun that sits in the specifier of TP, the noun with the least-marked case, etc.).*

todo

todo

2.3. Clarity issues

RC: *The major issue I ran into here was in the translations for many of the glossed examples. Words that are present in Turkish and the gloss line are absent from the translation (e.g. in (4), the word “instructor” is entirely missing) and the translations are imprecise (e.g. in (3), the word order is off). The authors should go through all of these carefully to ensure accuracy. This is critical to making sure the reader can have a clear understanding of these examples, even if they are not familiar with Turkish.*

I would also recommend careful editing for typos and awkward wordings. For example the paragraph on page 7 starting with “One example set of experimental items is in (5)...” feels out of place and repetitive.

We thank the reviewer for this comment. We corrected the errors in the translation and glossing. We also went over the text and edited instances of awkward wordings. All of the changes were marked with a red font in the manuscript.