Ashlyn Winship
LING 6477
Prof. Helena Aparicio
13 Dec. 2023

Mini Lab II
Replication experiment: Voice-mismatched VP ellipsis in Poppels & Kehler 2019

1 Introduction
This investigation is designed to replicate part of the research by Poppels and Kehler 2019, which investigates Mismatch Asymmetry. Their investigation suggests that a penalty against passive ellipsis clauses in subject focus environments may explain the phenomenon (2019:1).

According to the Recycling Hypothesis (RH) (Arregui et al. 2006), Mismatch Asymmetry is attributed to a "processing phenomenon" (Poppels & Kehler 2019:4) on the parts of both the speaker and the hearer. When the speaker is planning the next clause, they are more likely to remember a passive clause as an active one than vice versa, and then produce an active ellipsis clause. Similarly, the hearer is more likely to remember a passive clause as an active rather than vice versa, and then rate a [P->A] mismatch more favorably than an [A->P] mismatch (Poppels & Kehler 2019:3-4).

Poppels and Kehler conducted a series of experiments that focused on voice mismatch in VP-ellipsis, and investigated how well both syntactic and referential theories of Mismatch Asymmetry could account for the data they collected (2019:5). They concluded that "neither memory-based explanations such as the RH nor any ellipsis-independent explanation" are sufficient (2019:5). The first experiment found evidence that sentences in which the ellipsis clauses were in passive voice were rated significantly less favorably than those with active ellipsis clauses, regardless of whether the antecedent and ellipsis clauses were mismatched in voice (2019:7-8). The second experiment investigated whether that passive penalty is general or specific to ellipsis clauses by adding test items where the second clause did not feature ellipsis, and their results suggest that the penalty is specific to ellipsis clauses (2019:9-10). This paper is a replication of Poppels & Kehler's second experiment. Despite a small sample size, it found similar results. This supports the idea that there is a passive penalty that applies to ellipsis clauses that does not apply to clauses without ellipsis.

2 Experiment
This experiment is a replication of Poppels and Kehler's second 2019 experiment, and it is designed to test whether the previously observed passive penalty is general or ellipsis-specific.

2.1 Methods
2.1.1 Design and materials
The experimental stimuli and fillers were taken from the GitHub repository for Poppels and Kehler's 2019 experiment, so the participants in this study encountered the same materials as the participants in Poppels and Kehler's original experiment.

2.1.2 Procedure
The experiment was presented as a 5-point Likert scale acceptability judgment task administered through PC Ibex (demo). Each participant initially answered 3 training questions,

with feedback designed to teach them the difference between acceptability, plausibility, and unacceptability. Then, each participant encountered 24 experimental items, which were organized in a Latin square design, and 48 filler items. The order of the experimental items and fillers was randomized for each participant.

2.1.3 Participants

6 native speakers of English participated in the experiment, 3 males and 3 females. They ranged in age from 22 to 75,and all spoke Standard or Southern American English. Participants were not compensated.

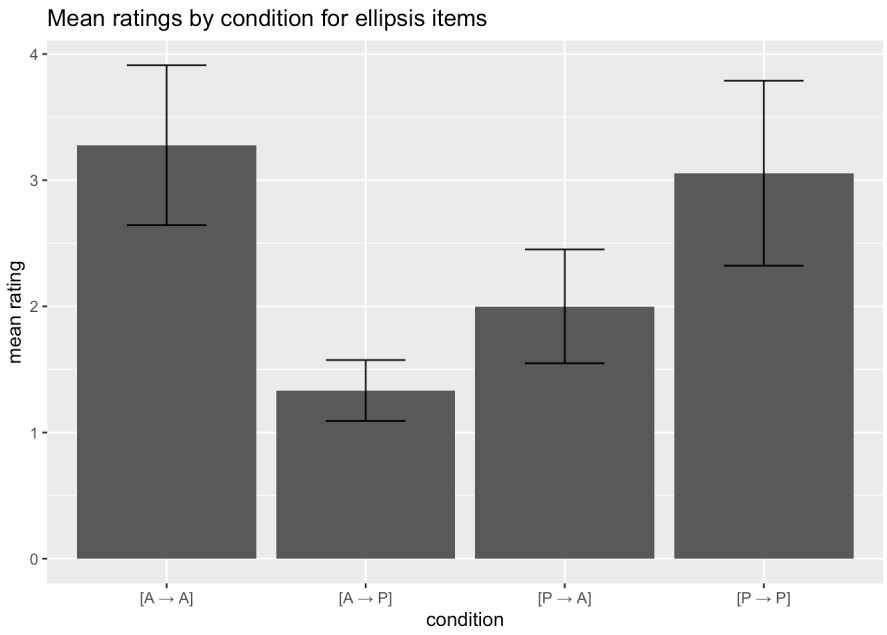
2.2 Predictions

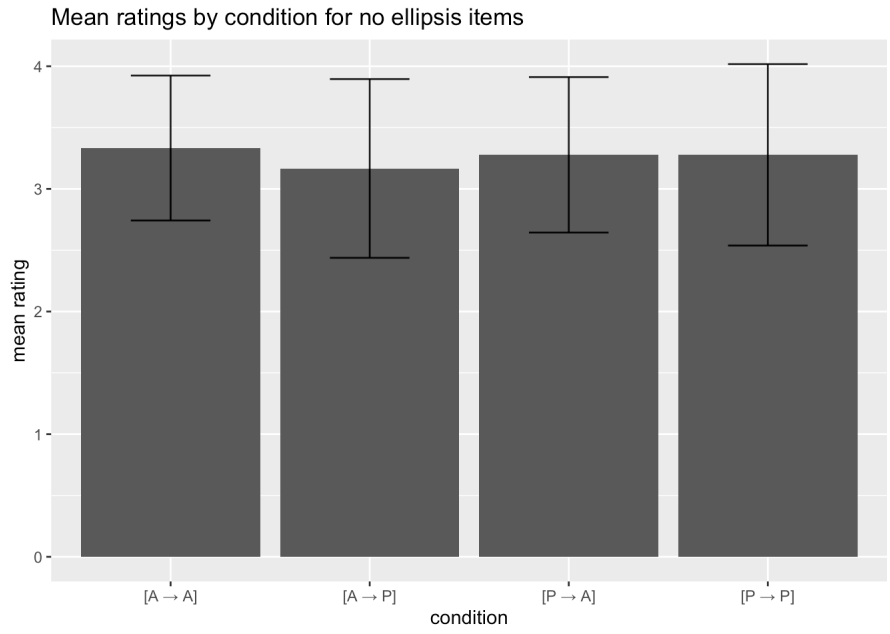
Poppels & Kehler performed mixed-effects regression analyses on their data, split by ellipsis and no-ellipsis items. These regressions showed both mismatch and passive penalties for the ellipsis condition, but no mismatch penalty and a marginal passive penalty for the no-ellipsis condition (2019:9-10). To replicate these results (without performing regression analyses), we expect to see significant differences between the ratings for [A->A] and [P->A] sentences, and between [P->P] and [A->P] sentences, for the ellipsis condition, and no significant differences between those sentence pairings for the no-ellipsis condition.

2.3 Results

None of the participants were discarded, so the data below comes from all 6 participants. The data was cleaned with a script provided by PC Ibex (source), and Google Sheets was used to filter that data down to the participants' responses for the experimental items and the fillers, and to organize the data according to the following columns and levels: **itemType** (filler, experimental), **response** (participant's rating, 1-5), **itemNo** (1-24 for experimental items, or NULL), **sid** (numeric vector, 1-n), **condition** ([A -> A], [P -> A], [P -> P], [A -> P], <NA>), **voice.cl1** (Active, Passive, <NA>), **voice.cl2** (Active, Passive, <NA>), **ellipsis** (TRUE, FALSE, <NA>), **fillerType** (upper-bound non-elliptical, upper-bound elliptical, lower-bound elliptical, lower-bound non-elliptical, <NA>), **mismatch** (TRUE, FALSE, <NA>), **voice.ellipsis** (passive, active, <NA>). The data was analyzed using RStudio.

With n=6, the total number of responses for experimental items was 144, and for fillers was 288 (432 tokens total). The mean rating for the filler items was 2.86. The graphs below show the mean ratings between sentence type for the ellipsis and no-ellipsis conditions, respectively.





Paired t-tests were used to analyze the data, because the data points were not independent (participants each provided multiple judgments).

Within the ellipsis sentences, there was a significant difference in the scores for [P->P] (M=3.056, SD=1.474) and [A->P] (M=1.333, SD=0.4851) conditions; $t(17)=4.2772$, $p<0.01$. There was also a significant difference in the scores for [A->A] (M=3.278, SD=1.274) and [P->A] (M= 2.000, SD=0.9075) conditions; $t(17)=3.1734$, $p<0.01$.

Within the no-ellipsis sentences, there was not a significant difference in the scores for [P->P] (M=3.278, SD=1.487) and [A->P] (M=3.167, SD=1.465) conditions; $t(17)=0.24338$,

p=0.8106. There was also no significant difference in the scores for [A->A] (M=3.333, SD=1.188) and [P->A] (M=3.278, SD=1.274) conditions; t(17)=0.23609, p=0.8162.

2.4 Discussion

Where Poppels & Kehler found a small passive penalty in the no-ellipsis sentences, this experiment found no such penalty. However, with a larger sample size, a similar penalty might have been observed. Regardless, Poppels & Kehler disregarded this effect.

Despite its small sample size, this experiment supported the hypotheses that the differences between ratings of [A->A] and [P->A] sentences, and between [P->P] and [A->P] sentences, would be significant for the ellipsis sentences, but not for the no-ellipsis sentences. This replicates the effects observed by Poppels & Kehler.

3. Conclusion

This small experiment replicates the effects found in Poppels & Kehler's second experiment, which tested whether the passive penalty observed on the ellipsis items in their first experiment was specific to those ellipsis items, or a general penalty. Their logistic regression found that both mismatch and passive effects significantly affected the ratings of sentences in the ellipsis test items, but not in the no-ellipsis test items. By comparing [A->A] and [P->A] sentences, and [P->P] and [A->P] sentences, I observed the similar results: These comparisons yielded significant differences in acceptability for ellipsis test items, but not for no-ellipsis test items. These results, along with Poppels & Kehler's, suggest that the passive voice in ellipsis constructions is generally dispreferred.

4 References

Poppels, T. & Kehler, A. (2019). Reconsidering asymmetries in voice-mismatched VP-ellipsis. *Glossa: A journal of general linguistics 4*(1), 1-22. DOI: https://doi.org/10.5334/gjgl.738.