Syntactic Effects on Agreement Attraction in Vocab-Limited Reading Experiments

Suhas Arehalli (Macalester College) & Tal Linzen (NYU)

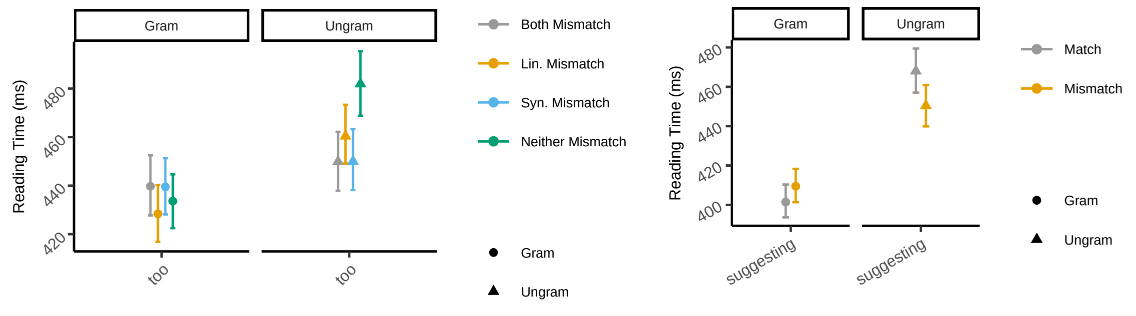
Speakers make subject-verb agreement errors more often in the presence of *attractors*: noun phrases that don't match the clause's subject in number [1]. Syntactic factors have been shown to modulate this effect (i.e., [2]). The number of proposed accounts for these agreement attraction effects highlights the need for precise computational models that can clearly determine which results each account can explain [3]. This can be challenging for a few reasons: (1) Probabilistic models that rely on learning from syntactically annotated data (i.e., models that parse like RNNGs [4]) struggle to model many existing experiments since their materials often contain words outside of the limited vocabularies of annotated corpora (i.e., [5]); (2) Many existing experiments have limited power and/or are scattered across different paradigms that require different linking hypotheses to simulate, which limits our ability to assign blame if a model doesn't simulate a particular result. We help address these issues by conducting conceptual replications of two English experiments that investigate syntactic factors on agreement attraction in self-paced reading (SPR) with materials that use a vocabulary consistent with the Penn Treebank [5]. We provide high-powered conceptual replications of prior work, as well as simulations using three NLM architectures (LSTMs [6], GPT-2 small [7], and RNNGs [4]).

Analysis: We exclude data from participants who had $< 80\%$ accuracy on comprehension questions. Human data is analyzed at spillover 1 using Bayesian linear mixed-effects models which assume an ex-Gaussian distribution of RTs to account for the right-tailed nature of reading times. Model surprisals are analyzed using linear mixed effects models.

E1: Syntactic Distance: We constructed 16 sentences like [2] with nested PP modifiers in the subject such that an attractor in the outer PP is hierarchically closer to the verb, but an attractor in the inner PP is linearly closer to the verb (Ex. 1). We analyzed data from 730 Prolific participants. We find an attraction effect from the syntactically closer attractor (95% CI: $[-23.38, -7.54]$) at the first spillover region, and a linear hypothesis test suggests the difference in attraction effects from the two attractors is significant (95% CI: $[-25.51, -7.88]$), replicating [2]. By contrast, NLM simulations show no difference between the effects from the 2 attractors (LSTM: $p = 0.87$, GPT2: $p = 0.22$, RNNG: $p = 0.13$).

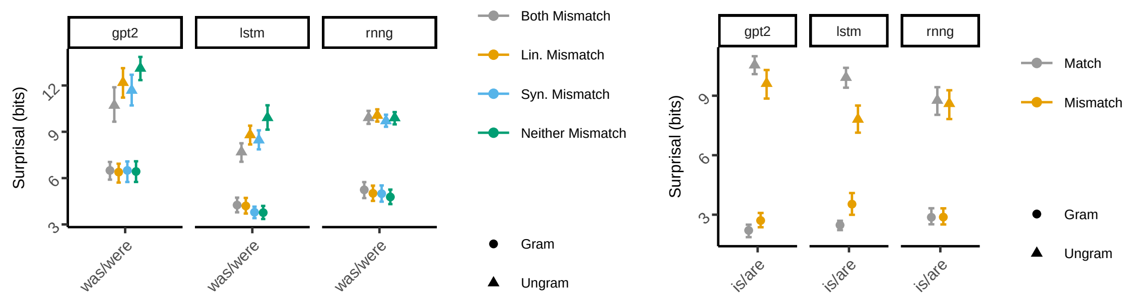
E2: Object Attraction: We constructed 16 sentences where attractors are syntactic objects that undergo movement in wh-questions (Ex. 2), a configuration that challenges structural accounts of agreement errors [8]. Extending a design from prior work ([8], which adapted [9] for SPR), we manipulate the attractor's number and the grammaticality of the sentence. We analyzed data from 470 Prolific participants. We find a significant agreement attraction effect (95% CI: $[-22.90, -8.61]$), replicating [8, 9]. NLMs other than the RNNG capture this effect (LSTM: $p < 0.001$, GPT2: $p < 0.01$, RNNG: $p = 0.79$).

Conclusions: We conceptually replicate two experiments [2, 8] that explore structural influences on agreement error patterns using materials with a Penn Treebank compatible vocabulary, with one pattern not emerging in any of our baseline NLMs. We see our data as a modeling target for more syntactically-aware quantitative models of agreement.



(a) (b)

Figure 1: Reading times at spillover 1 from the syntactic distance (1a) and object attraction (1b) experiments. Errorbars are 95% Confidence Intervals.



(a) (b)

Figure 2: Surprisals at the verb from the syntactic distance (2a) and object attraction (2b) simulations. Errorbars are 95% Confidence Intervals.

(1) The offer [for my share(s) [in the company(ies) $]_{PP}$ $]_{PP}$ quite frankly was/*were too low to consider.

(2) I know [[which stock(s) $]_1$ the analyst is/*are suggesting t_1 to the firm.]

1. Bock, K. & Miller, C. A. *Cog. Psych.* **23,** 45–93 (1991). 2. Franck, J., Vigliocco, G. & Nicol, J. *Language and Cognitive Processes* **17,** 371–404 (2002). 3. Yadav, H., Smith, G., Reich, S. & Vasishth, S. *JML* **129,** 104400 (2023). 4. Dyer, C., Kuncoro, A., Ballesteros, M. & Smith, N. A. in *NAACL* (2016), 199–209. 5. Marcus, M. P., Santorini, B. & Marcinkiewicz, M. A. *Comp. Ling.* **19,** 313–330 (1993). 6. Gulordava, K. et al. in *NAACL* (2018), 1195–1205. 7. Radford, A. et al. *Language Models Are Unsupervised Multitask Learners* tech. rep. (2019). 8. Arehalli, S., Linzen, T. & Legendre, G. in *AMLaP* (2021). 9. Dillon, B., Staub, A., Levy, J. & Clifton, C. *Language* **93,** 65–96 (2017).