The Levis and Vasishth model predicts acceptability judgments, but not RTs for attraction errors

Anna Laurinavichyute (University of Potsdam, National Research University Higher School of Economics), Titus von der Malsburg (University of Potsdam, MIT)
anna.laurinavichyute@uni-potsdam.de

The Lewis and Vasishth model of sentence processing (LV05, Lewis and Vasishth 2005) predicts the well-known morphosyntactic attraction effect in ungrammatical sentences: faster processing of the verb that mismatches the number of the subject, but matches the number of a non-subject (attractor) noun, as in *The difference between the studies are* (as compared to *The difference between the study are)*. The LV05 also predicts a similar effect in the semantic domain: faster processing of the verb that mismatches the semantic features of the subject, but matches those of the attractor, as in *The butler with the plate shattered* (compared to the *The butler with the tie shattered*). Note that this effect is expected in sentences with intact morphosyntax. This facilitation due to semantic match with the distractor – we will call it semantic attraction – was first demonstrated in reading by Cunnings and Sturt (2018), and replicated in an acceptability judgment task as well as directly compared to morphosyntactic attraction by Laurinavichyute and von der Malsburg (2020). We evaluate this surprising prediction of LV, which distinguishes is clearly from other accounts of attraction, on the latter dataset, and present a computational evaluation of the LV05 model's predictions on acceptability judgments and RTs.

Dataset. In a large-scale single-trial online experiment (N=2,498), participants memorized a verb presented in capitals, then pressed a button to see a sentence fragment, and decided whether the verb was a good continuation of the fragment. The verb never matched the *subject* noun (see Table 1): it mismatched either the subject's number (morphosyntactic violation), or meaning (semantic violation), or both (double violation). At the same time, the verb could mismatch or match the *attractor* noun in number (morphosyntactic attraction), meaning (semantic attraction), or both (double attraction). This set of conditions was used to test for agreement attraction, semantic attraction, and double attraction. Question response accuracies (Fig. 1) demonstrated similarly-sized effects of morphosyntactic and semantic attraction both in conditions with single violation and in conditions with double violation of subject-verb fit, but no interaction between the effects. The analyses of response times (Fig. 1) demonstrated slowdowns in judgment times for both morphosyntactic and double attraction, but not for semantic attraction effects.

Modeling. The main assumption of the LV05 model is that syntactic structure is built incrementally, and integration of a new constituent requires retrieval of the attachment site from memory. Retrieval operation relies on retrieval cues, such as +subj or +pl. If no constituent provides a perfect match for retrieval cues or if noise in the system decreases the activation of the matching constituent in memory, retrieval may fail and no structure will be built. We assumed that for our task, failure to retrieve a constituent and to build a structure leads to a rejection of the stimulus as ill-formed (correct response), while building any structure leads to acceptance of the stimulus (incorrect response). The resulting model fit for responses quantitatively captured six out of eight condition means (Fig. 1), and LV05 in general qualitatively predicted a decrease in accuracy due to every type of attraction. In contrast, model predictions for RTs mismatch the data in a systematic manner (see page 3 for details): across the whole range of plausible parameter values the model predicts either speedups or no difference for attraction conditions (Fig. 2), while we mainly observed slowdowns (Fig. 1). This mismatch between predicted speedups and observed slowdowns is not unique to the modeled dataset: slowdowns in judgment times are consistently observed for attraction effects (e.g. Staub 2009, Schlueter et al. 2019, Avetisyan et al. 2020), thus providing a systematic challenge for LV05. To accommodate the data, the model is likely to require an additional processing component that operates on top of structure building and specifically models processes deployed in the grammaticality judgment task.

Example 1:

Condition	Violation	Attraction
a. The drawer with the handle OPEN	morphosynt.	none
b. The drawer with the handles OPEN	morphosynt.	morphosynt.
c. The drawer with the handle CUTS	semantic	none
d. The drawer with the knife CUTS	semantic	semantic
e. The drawer with the handle CUT	double	none
f. The drawer with the knives CUT	double	double
g. The drawer with the knife CUT	double	semantic
h. The drawer with the handles CUT	double	morphosynt.

Note: [*b* vs. *a*] tests morphosyntactic attraction in single subject-verb fit violation; [*d* vs. *c*] tests semantic attraction in single subject-verb fit violation; [*f* vs. *e*] tests double attraction in double subject-verb fit violation; [*h* vs. *e*] tests morphosyntactic attraction in double subject-verb fit violation); [*g* vs. *e*] tests semantic attraction in double subject-verb fit violation.

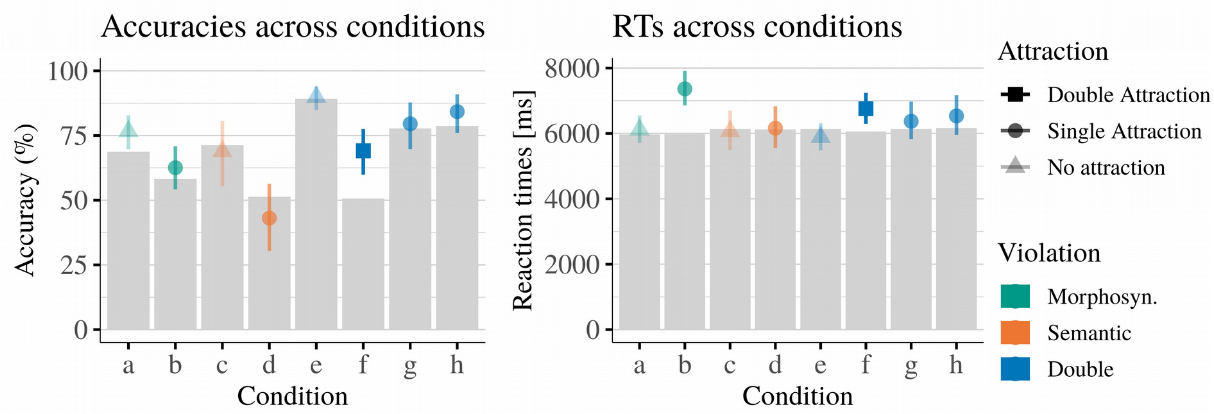

Figure 1. Estimated accuracy rates (left) and response times (right) with 95% credible intervals. Grey bars represent the predictions of the best-fitting set of parameters of LV05. Data means are captured if the top of the bar is contained within the 95% CrI.

References:

Lewis and Vasishth (2005) Cogn. Sci.
Cunnings and Sturt (2018), JML
Laurinavichyute and von der Malsburg (2020), doi: 10.31234/osf.io/hk9nc
Staub (2009) JML
Schlueter, Parker, & Lau (2019) FRONT PSY
Avetisyan, Lago, & Vasishth (2020) JML
Engelmann, Jäger, & Vasishth (2019) Cogn Sci., https://tinyurl.com/wo4z2ew

Modeling details. We modified the interACT implementation of LV05 (Engelmann et al. 2019) to model parsing with three instead of two retrieval cues: structural (+subj), morphosyntactic (+pl), and semantic (+can_cut). The critical parsing outcome that affected resulting judgment was whether the retrieval of the subject failed. We therefore varied two parameters that affected the probability of a retrieval failure: the level of noise in the system (between 0.05 and 0.45 in 5 steps of 0.1; the more noise, the less the outcome of a retrieval depends on retrieval cues matching the features of an item in memory) and the retrieval activation threshold (between -1.5 and 1.5 in 13 steps of 0.25; the higher the threshold, the lower the probability that any item is retrieved from memory). For modeling reaction times, we additionally varied the latency factor to scale model predictions into a numerical range comparable with the data (between 7.5 and 15 in 16 steps of 0.5). The simulation was run for 5000 iterations for each combination of parameters. Prediction error for both accuracy and RT modeling was quantified in terms of the average mean-squared error across the eight experimental conditions.

The slowdown that LV05 predicts for attraction conditions follows directly from the specification of the model and the mapping from modeling outcomes to acceptability judgments. Recall that retrieval failures are mapped onto correct responses, and compare the time it takes to register a retrieval failure to the time needed for a successful retrieval:

Retrieval failure = latency factor × $e^{-\tau}$
Successful retrieval = latency factor × e^{-A}

Here, A is the activation of the chunk that is retrieved, and τ is the retrieval activation threshold. For any chunk to be retrieved from memory, its activation A must be greater than τ, therefore, any retrieval will necessarily be faster than retrieval failure. It follows that control conditions without attraction with higher proportion of retrieval failures are predicted to be processed longer than conditions with attraction. The only exception is the configuration where the retrieval activation threshold τ is so high that retrievals fail in all conditions all the time — in that case, there is no difference in processing times between conditions. Importantly, there are no parameter configurations predicting a positive difference between the conditions with attraction and their respective control conditions without attraction, corresponding to the commonly observed slowdown.

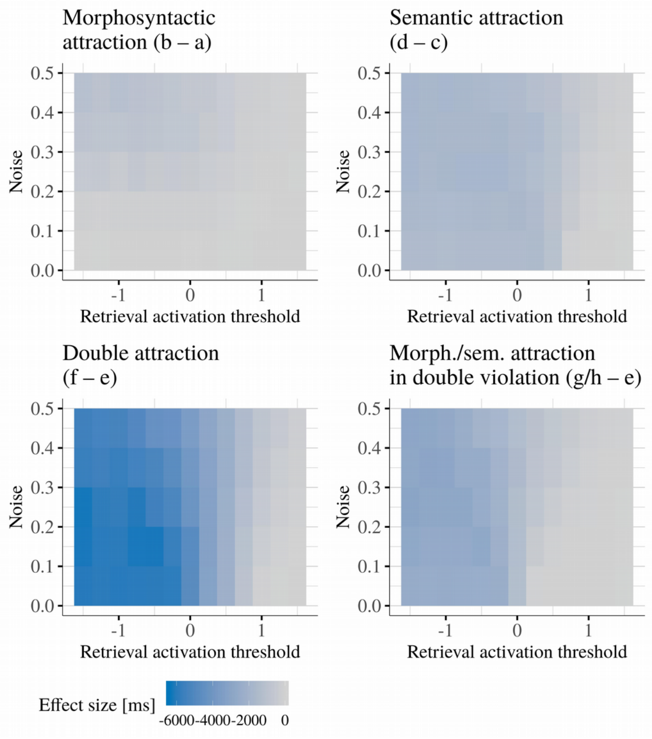

Figure 2. Predicted direction and magnitude of different attraction effects for the range of the varied parameter values: in reaction times, attraction is predicted to lead either to a speedup, or to no difference between conditions, but never to a slowdown.