Entertaining ungrammatical structures can succeed where surprisal struggles

Garrett Smith (Univ. of Connecticut) & Whitney Tabor (Univ. of Connecticut, Haskins Laboratories)
garrettsmith78@gmail.com

In some theories of sentence processing, parsing is strictly controlled by a symbolic grammar. The highly successful surprisal theory (Hale, 2001; Levy, 2008; Smith and Levy, 2013) has been implemented in this way, using a probabilistic context-free grammar to define the space of possible structures. At each word, the parser distributes probability over possible structures given the input; processing takes more time if this probability distribution has to be updated a lot after processing a new word. While surprisal theory has focused on comprehension, we apply it to a simple case of production, the verb selection task of Staub (2009): Subject noun phrases (NPs) are presented in rapid serial visual presentation (RSVP), followed by a two-alternative forced choice between a singular and plural verb form, followed by RSVP for the rest of the sentence. We can apply surprisal to this task by assuming the standard probability-updating mechanism for the RSVP parts. For the verb choice, we assume that the parser first updates its probability distribution so that only singular and plural verbs have non-zero probability. We assume this update is essentially cost-free in our stimuli since the verbs are very predictable. Then, it samples a verb from the distribution over the two verbs and updates its probability distribution again, to 100% on structures compatible with the chosen verb and 0% on the rest. The time it takes to produce the verb is essentially proportional to the second update time. One consequence of this approach is that, when the probability of singular and plural verbs is relatively evenly distributed (i.e., the distribution has high entropy), processing times are slower. In such a case, whichever verb is sampled will incur a large processing cost. When the distribution has lower entropy, the parser will usually sample the highest-probability verb, so the second update takes little time and processing is faster on average.

We reanalyzed the verb selection data from Smith, Franck, and Tabor (2018) to test this prediction. Smith et al. (2018) studied pseudopartitive subject NPs like those in Table 1. The response times (Fig. 2, left panel) are somewhat compatible with surprisal theory: Within the +N2 (second noun present) condition, faster verb production times are associated with lower entropies. However, across all conditions, some cases with similar entropy have very different response times (e.g., +N2 Collections and -N2 Measure Phrases). Thus, while the results are partially consistent with surprisal theory, the theory does not generate the full pattern of results.

An alternative theory, self-organized sentence processing (SOSP; Smith et al., 2018; Smith and Tabor, 2018; Kempen and Vosse, 1989), allows less-than-perfect structures to impact parsing, in contrast to current surprisal accounts. Indeed, the interaction between fully grammatical and imperfect structures is SOSP's key mechanism for explaining processing time effects (Smith and Tabor, 2018). In SOSP, lexically anchored syntactic treelets compete to form (sometimes imperfect) structures. The harmony of a structure—a measure of semantic and syntactic feature match (Smolensky, 1986)—determines how strongly the structure competes against other structures, with higher-harmony structures being more likely to form and faster to build (Smith and Tabor, 2018). We implemented a simple SOSP model of the competition between N1- and N2-headed structures for the pseudopartitive subject NPs. Experimentally derived semantic features from Smith et al. (2018) were used to calculate the harmonies of different structures. The model produced parse-formation times quite similar to the human data (Fig. 2, right panel), including the divergence between +N2 and -N2 conditions. This is because eliding N2 in the -N2 condition penalizes the harmony of attaching it as subject of the verb, thus slowing SOSP's structure building compared to the +N2 conditions. An interesting open question is whether Noisy Channel Suprisal (e.g., Futrell and Levy, 2017) which puts some probability on ungrammatical strings (but not structures) can fit these data better than standard surprisal. SOSP raises the alternative possibility that the difficulty of constructing ungrammatical structures may be a crucial determinant of processing times.

Table 1: Smith et al. (2018) materials; 64 test items and 64 fillers were used.; N = 57. Bracketed words were elided in the -N2 conditions. The first noun was always singular, and the second always plural.

N1 Type	Item
Containers	A basket [of flowers]
Collections	A pile [of flowers]
Measure Phrases	A lot [of flowers]
Quantifiers	Many [flowers]

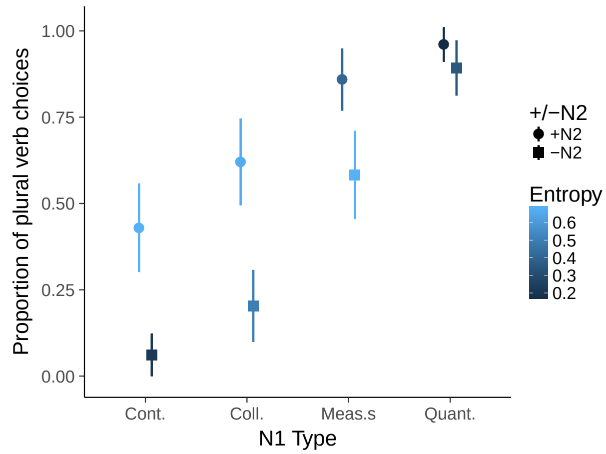

Figure 1: Smith et al. (2018) participant verb choices.

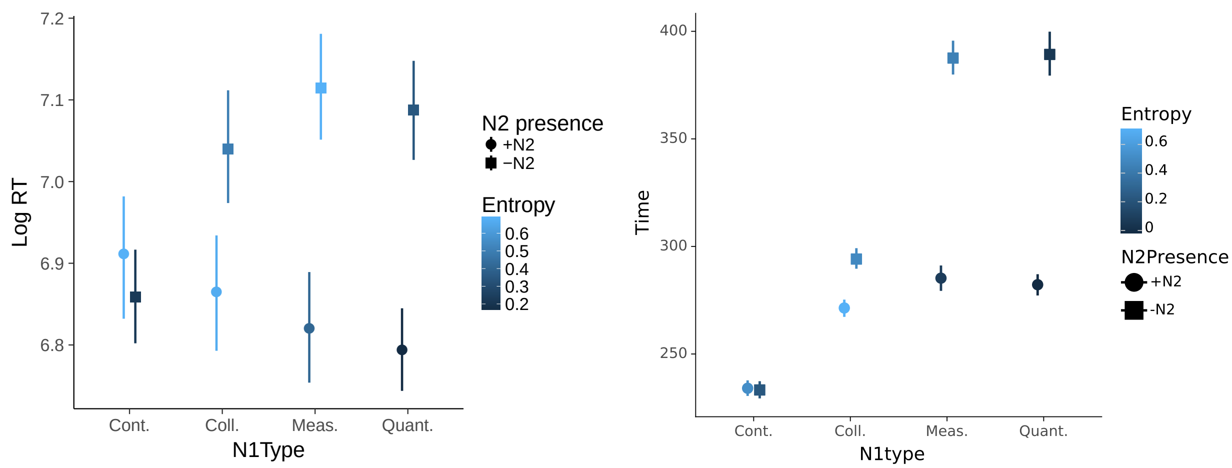

Figure 2: Human (left) and the SOSP model's (right) verb choice times (95% CIs). The model was run 2000 times per condition. A mixed effects analysis of human RTs (maximal random effects structure) revealed a main effect of $\pm$N2 (+N2 faster) and an interaction between $\pm$N2 and N1 Type, but no main effect of N1 Type.

J. T. Hale, in *NAACL* (2001), pp. 1–8.
R. Levy, Cognition **106**, 1126 (2008).
N. J. Smith and R. Levy, Cognition **128**, 302 (2013).
A. Staub, Journal of Memory and Language **60**, 308 (2009).
G. Smith, J. Franck, and W. Tabor, Cognitive Science **42**, 1043 (2018).
G. Smith and W. Tabor, in *Proceedings of the 16th International Conference on Cognitive Modeling*, edited by I. Juvina, J. Houpt, and C. Myers (2018), pp. 138–143.
G. Kempen and T. Vosse, Connection Science **1**, 273 (1989).
W. Tabor and S. Hutchins, Journal of Experimental Psychology: Learning, Memory, and Cognition **30**, 431 (2004).
P. Smolensky, in *PDP*, edited by Rumelhart, McClelland, and the PDP Research Group (MIT Press, 1986), vol. I: Foundations, pp. 194–281.
R. Futrell and R. Levy, in *EACL 15* (2017), vol. 1, pp. 688–698.