Lexical Effects in Structural Forgetting: Evidence for Experience-Based Accounts and a Neural Network Model

Michael Hahn (Stanford), Richard Futrell (UC Irvine), Edward Gibson (MIT)
mhahn2@stanford.edu

Doubly-nested center embedding structures give rise to **structural forgetting effects** whereby an ungrammatical sentence in which a verb is missing (Example 1b) may be perceived as grammatical, sometimes even more so than its grammatical variant (Example 1a) (Gibson & Thomas, 1999; Häussler & Bader 2015). Here, we explore a previously unexamined aspect of structural forgetting: the effects of *differences in lexical expectations* for a noun to be followed by a complement clause. Building on the theory of Futrell & Levy (2017), we account for the results using a neural network model and a rational model of noise in memory. **Prediction about Lexical Differences:** *If two nouns differ in the rates at which they are modified by a complement clause (e.g., "fact" vs. "report"), then the forgetting effect should be stronger for the noun where such modification is less frequent ("report")*. This prediction follows broadly from experience-based accounts (e.g. Frank et al., 2016). We test this prediction in materials where the top noun is modified by a sentential complement (2a-b), e.g. modification is more likely for "*fact*" than for "*report*". **Experiment 1: Production** We selected twelve nouns taking sentential complements, and constructed prefixes on the pattern of (3). We asked 144 participants to complete these prefixes to complete English sentences, and coded whether three verbs or fewer were present in the responses. In accordance with our prediction, the rate with which a verb was omitted was lower when P(that|the NOUN) was higher (Figure 1 and Table 1, e.g. 19% for "fact that" vs. 47% for "report that"). **Experiment 2: Ratings** We conducted a rating study comparing (2a) and (2b), using a pool of 35 nouns and 20 continuations. Each of 120 participants saw items constructed from 20 nouns randomly matched with these continuations. Half of the critical trials were presented in the (un)grammatical conditions (2a and 2b). Figure 2 and Table 2 show that, according to our prediction, the difference in ratings between grammatical and ungrammatical conditions is higher when P(that|the NOUN) is higher. **Model** To model these results, we constructed a neural network model of incremental processing with noisy memory representations. The model of Futrell & Levy (2017) assumes that memory representations are subject to stochastic noise that deletes some words of the past context. Incremental predictions are obtained by (1) inverting the noise model to compute *p(past | memory)*, and (2) obtaining $p(\text{next word}|\text{memory}) = \sum_{\text{past}} p(\text{next word}|\text{past})p(\text{past}|\text{memory})$, a standard noisy channel inference (Gibson et al., 2013). We parameterized *p(past | memory)* and *p(next word | past)* using recurrent neural networks estimated on a Wikipedia corpus (2 billion words). We compared the <u>uniform noise model</u> assumed by Futrell & Levy (2017) with a <u>rational noise model</u> parameterized by a neural network that estimates per-word forgetting probabilities, and is optimized for the tradeoff between increasing recovery quality *log p(past | memory)* and decreasing the number of remembered words. We used these models to obtain surprisals for grammatical (V2 and V1) and ungrammatical continuations (only V1) after observing sentences such as (2a) up to the end of V3, and predicted the surprisal difference from *log P(that|the NOUN)*. Using the uniform loss model, no effect was observed. However, when using the rational loss model, ungrammatical continuations were more strongly surprising for nouns with high values of *log P(that|the NOUN)*, in accordance with Experiments 1 and 2 (Figure 3). Thus a model of rational inference that is sensitive to what is easiest to reconstruct explains structural forgetting effects better than previous models. Such a model also has the benefit of explaining why noisy channel models are particularly sensitive to loss of function words (as assumed by Levy et al., 2009; Gibson et al., 2013, without explanation): because they are most recoverable.

Examples

(1a) The patient who the nurse who the clinic had hired$_{V3}$ admitted$_{V2}$ met Jack$_{V1}$.
(1b) *The patient who the nurse who the clinic had hired$_{V3}$ met Jack$_{V1}$.
(2a) The fact / report that the student who the professor disliked$_{V3}$ was dropping the class$_{V2}$ made the professor happy$_{V1}$.
(2b) *The fact / report that the student who the professor disliked$_{V3}$ made the professor happy$_{V1}$.
(3) The fact that the student who the professor…

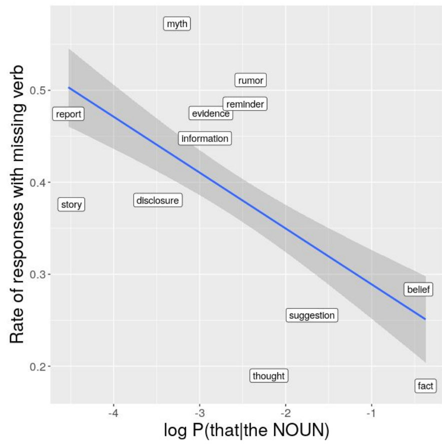

Figure 1: Rate of completions with missing verb, as a function of a corpus estimate of log P(that|the NOUN). Nouns for which this probability is high (e.g., *fact*) show lower rates of verb omission.

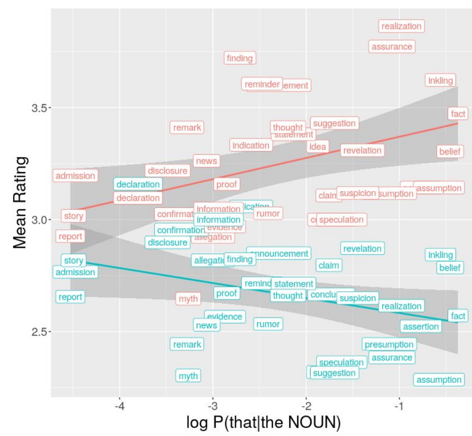

Figure 2: Average complexity ratings for grammatical (blue, 2a) and ungrammatical (red, 2b) sentences. Following Gibson & Thomas (1999), *higher ratings* indicate that a sentence is *harder to understand*.

	beta	SE	z
Intercept	-2.010	0.612	-3.28
log P(that|the NOUN)	-0.477	0.202	-2.35

Table 1: Logistic mixed-effects model predicting whether a verb was dropped in a completion.

	beta	SE	t
Intercept	2.955	0.070	42.30
grammatical (+/- 0.5)	-0.544	0.089	-6.12
log P(that|the NOUN)	0.015	0.019	0.79
Interaction	-0.148	0.036	-4.13

Table 2: Mixed-effects model of complexity ratings, with random effects for participants, nouns, and continuations. Similar results were obtained with a cumulative ordinal regression.

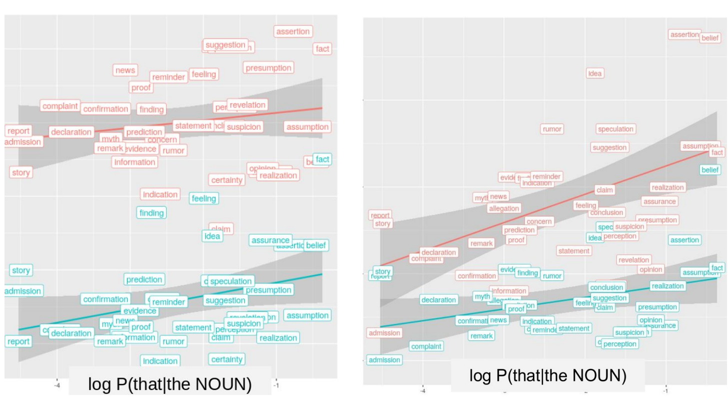

Figure 3: Predicted Surprisal on grammatical (blue) and ungrammatical(red) continuations, as a function of log P(that | the NOUN), for the uniform (left) and rational (right) noise models.