Verb omission errors: Evidence of rational noisy-channel language processing

Leon Bergen (Massachusetts Institute of Technology), Roger Levy (University of California, San Diego), & Edward Gibson (Massachusetts Institute of Technology)

bergen@mit.edu

Rational analysis; Noisy-channel models; Sentence production; Self-paced reading; English

Recent work by [1,2] suggests that human sentence processing maintains uncertainty in input representations and reconciles it with prior grammatical knowledge. Here we explore the prediction of such rational noisy-channel theories that morphosyntactic production errors and comprehension garden-pathing should be inducible when context strongly biases expectations about syntactic category.

We use unambiguous NN-phrases ("The voter hope") differing in a single morpheme from unambiguous NV-phrases ("The voter hoped"). A noisy-channel model where morphemic realization errors are possible makes four key predictions regarding the processing of such phrases. First, the Bayesian *size principle* [3,4] predicts that producers should complete NN-preambles as NV-preambles, creating apparent *verb omission errors* (1), more often than the reverse error. The NN-preamble can be produced from the NV-preamble by deleting a morpheme, whereas producing the NV-preamble from the NN-preamble requires inserting a morpheme. Because there are more ways to choose a random morpheme to insert than to delete, the posterior probability of NV given NN must be higher than NN given NV. Second, verb omission error rate should be higher the greater the NV bias of the phrase. Third, repetition errors for the preambles should pattern with other production errors: adding past or present tense morphology to NN-preambles should be more common than removing it from NV-preambles. Fourth, in comprehension, it should be possible to garden-path readers of "unambiguous" NN-preambles onto NV syntactic interpretations inconsistent with veridical wordforms, and see the behavioral hallmarks of garden-path disambiguation further downstream.

Experiment 1 evaluated the first three claims. 60 participants were given a visual presentation of the preamble for 1.5sec and had 13sec to retype the preamble and complete it (subject-verb agreement error patterns from [5] were separately replicated using this paradigm). 16 NV-biased (e.g. "The voter hope") and 16 NN-biased (e.g. "The almond roll") preambles were show to participants (bias was measured by completions for the ambiguous NN/NV preamble, e.g., "The voter hopes"). Restricting to completions without repetition errors, participants were significantly more likely to complete NN-preambles as NV-sentences (see 1) than vice-versa (27 vs. 1; $p<0.001$; ANOVAs with subjects and items as random factors). All but one of the NN-errors occurred on an NV-biased item. Finally, morphological errors in repetition occurred significantly more often for NN-preambles (12 vs. 1; $p<0.01$).

Experiment 2 evaluated the fourth claim using self-paced reading (120 participants, 16 items, 32 fillers).

Participants read sentences which contained a modifier-noun or NV-preamble as in Experiment 1; in critical but not in control conditions, NV and NN preambles were single-morpheme-edit neighbors (2). As a result, participants should be more likely to be garden-pathed in the critical condition. At the disambiguating word ("seemed") we obtained a significant interaction ($p<0.05$; lmer with random slopes for participants and items), with reading times superadditively high in the critical modifier-noun condition.

Together, these experiments constitute the first results indicating that comprehenders can be biased to morphosyntactic processing inconsistent with veridical input due to context-specific grammatical prior expectations for syntactic category.

(1) NN/NV: The voter hope(d)…

> **Example verb omission error:** The voter hope that the president will win.
> **Example verb insertion error:** The voter hoped was that the president will win.

(2) Critical modifier-noun: The voter hope for a victory seemed naïve.

> **Critical NV:** The voter hoped for a victory but seemed naïve.
> **Control modifier-noun:** The national hope for a victory seemed naïve.
> **Control NV:** Some voters hoped for a victory but seemed naïve.

References

[1] Levy, R. (2008). EMNLP. [2] Levy et al. (2009). *PNAS*. [3] MacKay, D. (2003). CUP. [4] Xu and Tenenbaum (2007) *Psychological Review*. [5] Bock & Miller (1991) *Cognitive Psychology*.