Can garden path effects be reduced to predictability?

Unpredictable words are read more slowly than predictable ones. It has been suggested that processing difficulty in the disambiguating region of garden path sentences can be reduced to predictability (surprisal; [3]), obviating the need for syntactic repair mechanisms [1]. The goal of this work is to test this hypothesis using predictability estimates from a recurrent neural network (RNN) language model [2]. We examine three types of garden path sentences:

NP/S: Mary saw (that) the doctor ⌈had been⌉ drinking.
NP/Z: When Mary visited (,) the doctor ⌈had been⌉ drinking.
MV/RR: The soldiers (who were) warned about the doctor ⌈conducted the late⌉ raid.

When the disambiguating context (in parentheses) is omitted, each of these constructions lead readers to assign a high probability to a parse that later on turns out to be incorrect: the critical region (in square brackets) is incompatible with that parse, producing slower reading times (RTs) compared with a version of the sentence that includes the disambiguating context.

Does the RNN model the syntax of garden path sentences? We looked at the parts-of-speech predicted by the model at the first word of each critical region (Figure 1).[1] The unambiguous conditions generally caused the model to correctly predict a verb, while all three ambiguous conditions led the model to instead predict a reasonable incorrect alternative prediction (e.g., a sentence boundary). These results show that the model is sensitive to the syntax of these constructions.

Surprisal is a better predictor in NP/S than in NP/Z or MV/RR: The surprisal of a word influences the RT of that word as well as subsequent words. [5] regressed the model surprisal of each word in a four-word window against RTs to determine the influence of surprisal over time.[2] We used their regression slopes to predict the RT of each word in our experiment with a weighted sum of the surprisals of the preceding words.[3] The mean unambiguous critical RT was subtracted from the corresponding mean ambiguous RT to obtain a predicted garden path effect (Figure 2). The model predictions were compared with the self-paced RTs reported by [4] (details in Table 1). The model's NP/S prediction was not significantly different from the human response ($p = 0.12$). The NP/Z and MV/RR responses predicted from the model's surprisal, while in the correct direction, were significantly smaller than the corresponding human responses (both $p \leq 0.001$).[4]

No evidence for a difference in the distribution of RTs between NP/S and fillers: We then examined the distribution of each construction's critical RTs when the context is ambiguous (i.e. enables a garden path effect) compared with the distribution of RTs on the same sentence positions in the filler sentences read by the same participants. Two-sample Kolmogorov-Smirnov tests revealed that the distribution of NP/S RTs did not significantly differ from the filler RT distribution ($p = 0.20$), while the distributions of NP/Z and MV/RR RTs did significantly differ from the distribution of their respective filler RTs (both $p < 0.001$).

Conclusions: These results suggest that processing difficulty in NP/S sentences may indeed be reducible to a predictability effect which is broadly present in reading (including in filler sentences). However, predictability seems insufficient to model NP/Z and MV/RR processing, and the distribution of RTs in these constructions differs from non-garden path RTs. These results suggest that NP/Z and MV/RR constructions may involve qualitatively different processing than NP/S constructions and general purpose reading.

---

[1] The model's lexical probabilities were mapped to the most likely corresponding part-of-speech.

[2] Their models also controlled for word frequencies and word lengths in that window as well as sentence position.

[3] The model RT prediction at a given word w_i is calculated based on the surprisal (S) of previous words:
RT$(w_i) = 0.84 \cdot S(w_{i-3}) + 0.92 \cdot S(w_{i-2}) + 1.53 \cdot S(w_{i-1}) + 0.53 \cdot S(w_i)$

[4] We also replicated this finding using a syntactic parser [6] to ensure results weren't driven by RNN syntactic deficits.

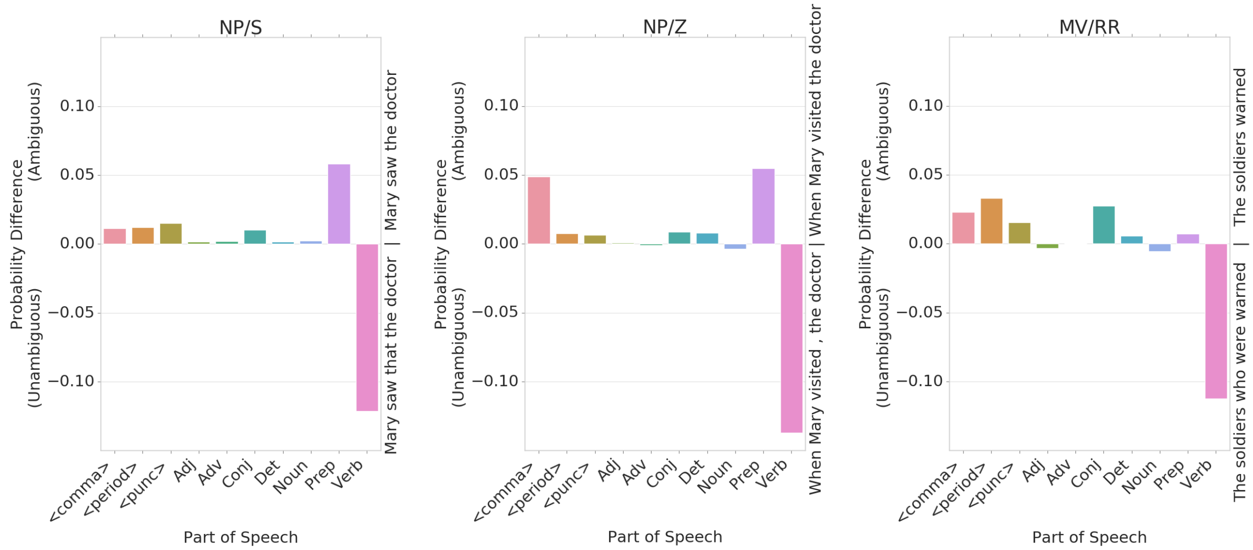

Figure 1: RNN part-of-speech probabilities at the first word of the critical region in the ambiguous condition (positive) minus those in the unambiguous condition (negative) in each construction.

	# Items	# Fillers	# Subjects
NP/S	20	80	224
NP/Z	20	80	224
MV/RR	40	80	73

Table 1: Details of self-paced reading data collected by [4].

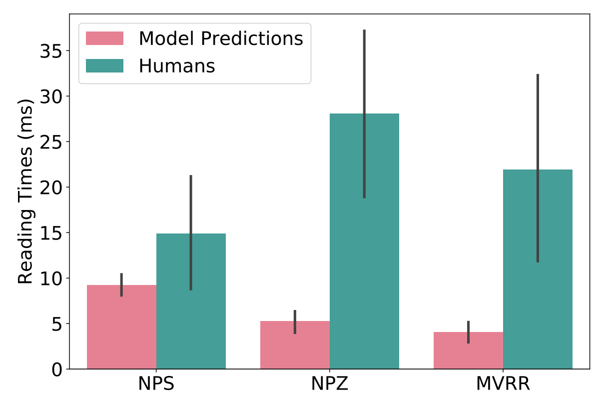

Figure 2: Model RT predictions and human RTs in critical regions (ambiguous RTs minus unambiguous RTs).

References

[1] Janet Fodor and Fernanda Ferreira. *Reanalysis in Sentence Processing*. Springer, Netherlands, 1998.

[2] Kristina Gulordava, Piotr Bojanowski, Edouard Grave, Tal Linzen, and Marco Baroni. Colorless green recurrent networks dream hierarchically. In *Proceedings of NAACL 2018*, 2018.

[3] John Hale. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American chapter of the Association for Computational Linguistics*, pages 159–166, Pittsburgh, PA, 2001.

[4] Grusha Prasad and Tal Linzen. Reassessing syntactic adaptation. Technical report, Cognitive Science, Johns Hopkins University, 2019.

[5] Nathaniel J. Smith and Roger Levy. The effect of word predictability on reading time is logarithmic. *Cognition*, 128:302–319, 2013.

[6] Marten van Schijndel, Andy Exley, and William Schuler. A model of language processing as hierarchic sequential prediction. *Topics in Cognitive Science*, 5(3):522–540, 2013.