Recurrent neural networks use discourse context in human-like garden path alleviation

Forrest Davis & Marten van Schijndel (Cornell University)
fd252@cornell.edu

Garden path sentences highlight the role of preceding linguistic context in parsing. Compare the following sentence (from [2]) with its unambiguous form: *The horse raced past the barn fell* vs. *The horse that was raced past the barn fell*. Readers experience confusion (known as a garden path effect; GPE) at the verb *fell* in the first sentence, having expected *raced* to be a main verb rather than the verb in a reduced relative clause (RC). The GPE is alleviated when *raced* is in an overt RC, as in the unambiguous second sentence. Recurrent neural network (RNN) language models (LMs) assign lower probability to *fell* in the reduced version compared to the overt version (analogous to a GPE; [3, 6]). Human GPEs can be alleviated by changing the discourse context to accommodate conversational implicatures of the RC [1, 4, 5]. In this work, we simulate [4] and [5] with RNNs, showing that similar discourse factors mitigate RNN GPEs, providing stronger empirical support for RNN use in modeling human behavior.

(1) Referentially supporting stimulus from [4]
- (a) **Context**:
  - (i) 1 knight - A knight and his squire were attacking a dragon. With its breath of fire, the dragon killed the knight but not the squire.
  - (ii) 2 knights - Two knights were attacking a dragon. With its breath of fire, the dragon killed one of the knights but not the other.
- (b) **Target**:
  - (i) Reduced - The knight killed by the dragon fell to the ground with a thud.
  - (ii) Unreduced - The knight who was killed by the dragon fell to the ground with a thud.

The RCs in (1b) produce a conversational implicature that there exists more than one *knight* who is distinguished by the modifier *killed by the dragon.* [4] demonstrated that the GPE from reading (1b-i) is alleviated when preceded by a context that contains multiple referents (1a-ii), in contrast with a context containing a single referent (1a-i). [5] demonstrated a similar alleviation for future tense contexts (e.g., (2)).[1] We trained 10 RNN LMs on 80 million words of Wikipedia text; 5 were trained on data shuffled by sentence to remove discourse cues during training (as in the widely used model from [7]), and 5 were trained on sentences in their original order. Information-theoretic surprisal $Surp(w_i) = -log\ p(w_i|w_1 ... w_{i-1})$ was calculated at the disambiguating region (verb+by as in [4] and [5]) of the stimuli in [4] and [5].

Results: 1) All 10 models exhibited significantly[2] more surprisal in the reduced condition than the unreduced variants (i.e. reduced RC sentences produced GPEs). **2)** The difference in surprisal between neutral contexts (e.g., (1a-i)) and alleviating contexts (e.g. (1a-ii)) differed significantly from zero only for models trained on ordered data, confirming that training on discourse context is crucial for exhibiting multi-sentence effects. **3)** In line with the results for humans, alleviating contexts **reduced** the surprisal at the disambiguating region compared to neutral contexts (Fig. 1 and Fig. 2). Our results show that RNN LMs trained on data containing discourse relations can handle multi-sentence phenomena similar to humans, suggesting that psycholinguists can use RNN LMs to model discourse impacts on sentence processing.

[1] Future contexts provide evidence that main verbs will not be past tense, so the past tense main verb interpretation becomes less preferred than the relative clause interpretation.
[2] After Bonferroni correction for 14 statistical tests, our significance threshold was 0.003. One sample t-tests tested whether each experimental condition mean differed from zero, and two sample t-tests tested whether there was a difference between model types (ordered vs shuffled).

(2) Temporally supporting stimulus from [5]

(c) **Context**:

(i) Past - Several students were sitting together taking an exam in a large lecture hall earlier today. A proctor noticed one of the students cheating.

(ii) Future - Several students will be sitting together taking an exam in a large lecture hall later today. A proctor will notice one of the students cheating.

(d) **Target**:

(i) Reduced - The student spotted by the proctor received/will receive a warning

(ii) Unreduced - The student who was spotted by the proctor received/will receive a warning.

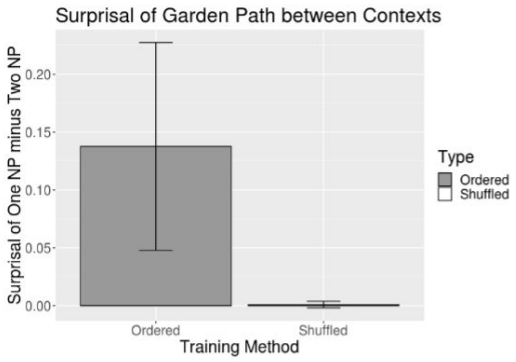

Figure 1: Differences between RNN LM surprisals at the verb+by region when preceded by a context with one referent (garden-path supporting) and when preceded by a context with two referents (garden-path alleviating). Stimuli are from Spivey-Knowlton et al. (1993).

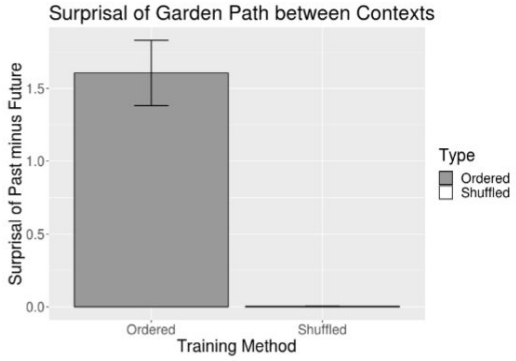

Figure 2: Differences between RNN LM surprisals at the verb+by region when preceded by a past context (garden-path supporting) and when preceded by a future context (garden-path alleviating). Stimuli are from Trueswell and Tanenhaus (1991).

References:

[1] Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30(3)*, 191-238.

[2] Bever, T. G. (1970). The cognitive basis for linguistic structures. *Cognition and the development of language, 279(362)*, 1-61.

[3] Frank, S. L., & Hoeks, J. (2019). The interaction between structure and meaning in sentence comprehension: Recurrent neural networks and reading times. P*syArXivpreprint:10.31234.*

[4] Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology, 47(2),* 276.

[5] Trueswell, J. C., & Tanenhaus, M. K. (1991). Tense, temporal context and syntactic ambiguity resolution. *Language and Cognitive Processes, 6(4)*, 303–338

[6] Van Schijndel, M., & Linzen, T. (2018). Modeling garden path effects without explicit hierarchical syntax. In *CogSci*.

[7] Gulordava, K., Bojanowski, P., Grave, E., Linzen, T., & Baroni, M. (2018). Colorless green recurrent networks dream hierarchically. In *NAACL*.