# A random walk down the garden path: A new implementation of self-organized parsing

Garrett Smith (Universität Potsdam)
gasmith@uni-potsdam.de

Models of human sentence comprehension typically assume that the parses people build during word-by-word language understanding are globally consistent with the grammar of their language: Only structures that follow all the rules of the grammar are considered as (partial) parses of a string of words. These models have been widely successful in explaining how people parse sentences (Levy, 2008; Lewis & Vasishth, 2005). However, local coherence effects, where locally viable but globally ungrammatical structures seem to compete with grammatical ones, present a challenge for traditional theories of human sentence comprehension (Tabor, Galantucci, & Richardson, 2004). An alternative theory, based on principles of self-organization, can explain these effects in a natural way. Under self-organization, words assemble themselves into larger syntactic structures according to violable constraints via purely local interactions. There is no global consistency checking; nevertheless, grammatical parses typically emerge on their own. Despite the theoretical innovation of previous self-organizing models, their implementations have suffered from opaque mathematical formalisms and limited coverage of empirical phenomena (e.g., Kempen & Vosse, 1989; Smith & Tabor, 2018). We present a new implementation, called *mparse*, that shows promise for overcoming these issues and making self-organization a more broadly and easily testable theory.

Mparse applies the master equation—used in chemistry and physics to describe continuous-time, discrete-state random walks (Oppenheim, Shuler, & Weiss, 1977)—to model human sentence comprehension. This is how it works: At each word in a sentence, mparse enumerates all possible partial and complete parses that are possible given the words so far and a grammar of binary dependency relations between words. These parse states include both merely locally viable structures and globally grammatically ones. The model jumps stochastically between parse states that differ only by a single dependency link (nearest neighbors), with jumps to more well-formed states being more probable than jumps to less well-formed states. A noise parameter controls how much the model prefers well-formed states over ill-formed ones (low noise = strong preference for well-formed states, high-noise = well- and ill-formed states treated equally). Well-formedness is penalized if a state has too few dependency links, includes longer dependencies, and/or includes word order violations. Reading times are modeled as the amount of time it takes mparse to reach a state with the maximum possible dependency links for the number of words so far (up to $w - 1$ links for $w$ words). Once mparse reaches such a state, it stops processing the current word, inputs the next word, adds new states based on the syntactic affordances of the new word, and resumes the random walk among the states. The master equation formalism offers powerful tools for understanding incremental sentence parsing and making detailed, quantitative predictions. Importantly, mean reading times for each word in a sentence can be calculated easily.

We tested mparse on local coherence effects (1) and the contrast between two types of garden path effects (2) in English (Sturt, Pickering, & Crocker, 1999). As shown in Fig. 1 (left), mparse correctly produces disproportionately slow mean reading times for *. . . at the player tossed. . .* from Tabor et al. (2004). It also correctly produces larger garden path effects (ambiguous - unambiguous) for NP/Z materials than NP/S materials (Fig. 1, right, Sturt et al., 1999).

These results demonstrate that this implementation of self-organization produces reading time predictions in line with existing experimental results. The proof-of-concept results presented here, though, barely scratch the surface of the information that can be gleaned from mparse's mathematical formalism. Future work will explore how the mathematical theory behind mparse can drive new empirical work. Work is also underway for extracting mparse's grammar from large, parsed corpora, opening the door to truly broad-coverage comparisons with competing models like surprisal (Levy, 2008) and cue-based retrieval (Lewis & Vasishth, 2005).

(1) The coach smiled at the player [who was] [tossed/thrown] the frisbee.

(2) a. NP/S: The woman saw [that] the doctor had been drinking.
b. NP/Z: Before the woman visited[,] the doctor had been drinking.

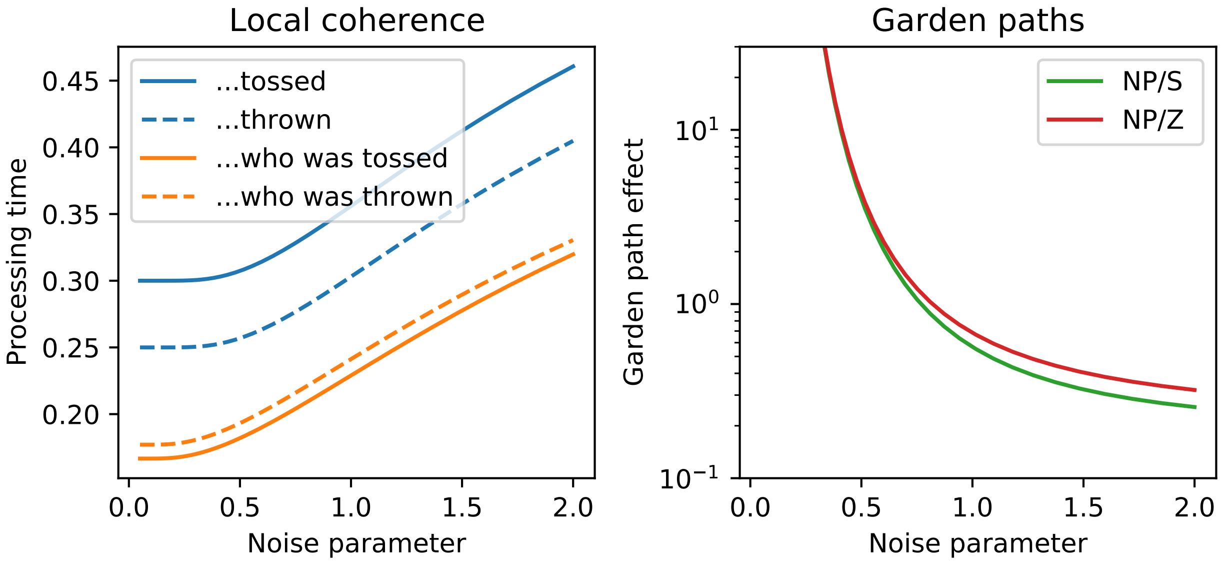

Figure 1: Mean processing times (arbitrary units) at *tossed/thrown* in (1) (left) and mean garden path effects at *had* in (2) (right). The garden path effects are the difference between the ambiguous and unambiguous conditions in (2). Note the logarithmic y-axis in the right panel. As the noise level decreases, the size of both garden path effects explodes because the probability of jumping from a relatively well-formed garden-path state to an ill-formed state intermediate between the garden path and the correct parse decreases rapidly, making repairing the garden path nearly impossible.

## References

Kempen, G., & Vosse, T. (1989). Incremental syntactic tree formation in human sentence processing: A cognitive architecture based on activation decay and simulated annealing. *Connection Science*, *1*(3), 273–290.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 375–419.

Oppenheim, I., Shuler, K. E., & Weiss, G. (1977). *Stochastic processes in chemical physics: The master equation*. MIT Press.

Smith, G., & Tabor, W. (2018). Toward a theory of timing effects in self-organized sentence processing. In I. Juvina, J. Houpt, & C. Myers (Eds.), *Proceedings of the 16th International Conference on Cognitive Modeling* (pp. 138–143). Madison, Wisconsin: University of Wisconsin.

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*, *40*, 136–150.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*(4), 355–370.