Syntax in Flux: Structural Priming Maintains Probabilistic Representations

Neal E. Snider (*nsnider@*) T. Florian Jaeger (*fjaeger@bcs.rochester.edu*)

Department of Brain & Cognitive Sciences, University of Rochester

Abstract

Much recent work suggests that language users employ probabilistic information during online production [3,16] and comprehension [9,18]. Probabilistic information is acquired early in life [2], and speakers are able to adapt their expectations to speakers and environments [14,7]. This suggests that speakers update probabilistic representations (cf. connectionist and other probabilistic approaches [17,15]). **We present evidence that speakers are indeed updating syntactic probability distributions.** The evidence comes from syntactic priming, which has been linked to implicit learning [6] and to surprisal-sensitive processing [12].

Prime Surprisal

- We propose that the language processing system is set up in such a way that, whenever an instance of a structure is processed, it is seen as a piece of evidence that affects the structure's probability distribution
- Maintenance of probability distributions is assumed to be an inherent part of the language processing system.
- Less probable syntactic structures, if observed, lead to a bigger change in the probability distribution, which in turn leads to an increased probability of reusing the same structure (cf. [10] on perceptual persistence; [11] on skill maintenance).

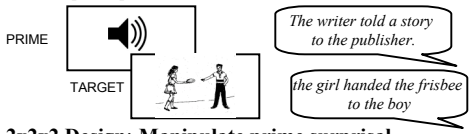
→More surprising primes should prime more strongly (i.e lead to a bigger increase in the probability of repetition)

$$surprisal(x) = \frac{1}{\log probability(x)}$$

- Surprisal-sensitive priming would also explain the well-known anti-frequency effect (the less frequent structure primes more, [5,8, i.a.])

Data: **Production priming experiment**

(45 participants, 24 items, 72 fillers).



2x2x2 Design: **Manipulate prime surprisal**

A-priori experience of prime x **Prime** structure

PO: The writer told / passed a story to the publisher

DO: The writer told / passed the publisher a story

DO-bias PO-bias

= higher surprisal

(estimated from norming expt.- 41 subjects)

x **Recent** experience:

2 blocks of 12 *vs.* 24 alternating primes

(counterbalanced)

Study 1: Surprisal-sensitive priming based on *a-priori* and recent distributions

Analysis: Mixed logit models
Collinearity: no fixed effect correlations > .25

Control Factors:
Participant intercept (random)
Item intercept (random)
Target verb bias (random)

Results:

Prime structure ($p > .5$)

Prime structure *X* ***a-priori*** **experience** ($\chi^2(1)=6.6$, $p < .05$)

The more surprising the prime structure given experience before the experiment, the more likely it will be repeated. This effect holds beyond the specific prime structure (DO vs. PO) and target bias.

→ *Evidence for surprisal-sensitive priming (previous experiments confounded prime and target surprisal,* [4])

Recent experience ($\chi^2(1)=4.2$, $p < .05$)

The higher the proportion of a prime structure in the preceding trials (excluding the most recent prime), the more likely speakers will produce that structure

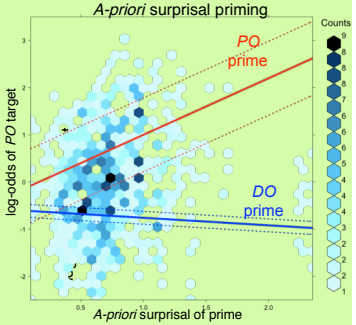
→*Cumulative priming* [13,12]

Prime structure X **Recent experience** ($\chi^2(1)=5.4$, $p < .05$)

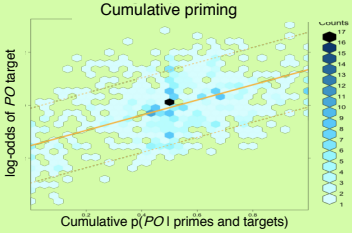
The more surprising the prime structure given experience during the experiment (primes), the more likely it will be repeated.

→ ***Evidence that speakers update their probability distributions over syntactic structures: prime surprisal is affected by both a-priori and recent experience***

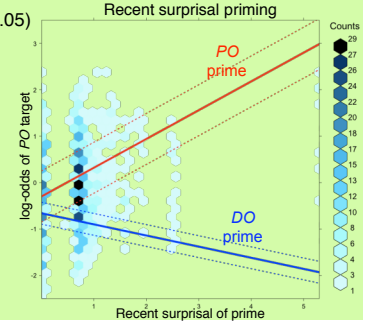
A-priori surprisal priming



Cumulative priming



Recent surprisal priming



Study 2: Modeling probabilistic cue combination in priming

We investigated how *a-priori* and recent experience are may be combined to produce the observed surprisal-sensitivity.

We systematically explored the parameter space of a priming model, incorporating surprisal-based activation-gain of the prime structure.

We assumed that :

(1) The overall surprisal for a PO structure at each trial *i* is a linear interpolation between *a-priori* and recent experience:

$$p(PO)_i = \omega * p(PO \mid norm_i) + 1 - \omega * p(PO \mid trials < i)$$

(2) The activation of a *PO* structure at trial *i* is positively correlated with the overall surprisal of PO:

$$Act(PO)_i = \begin{cases} \text{if } i{=}PO & \alpha * -log\ p(PO)_i \\ \text{if } i{=}DO & -\alpha * -log\ p(PO)_i \end{cases}$$

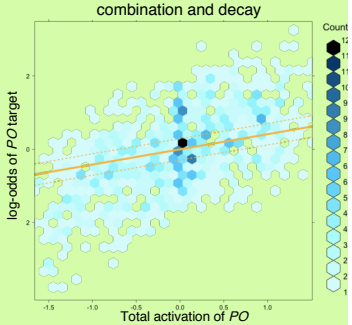
(equivalent to Prime x Surprisal interaction)

(3) Activation is subject to power-law decay [1], so that the total activation of the *PO* structure during a target trial will be:

$$TotalAct(PO)_{target} = \sum_{i=1\ldots target} Act(PO)_i \cdot (target - i)^{-\delta}$$

We varied ω, α, *and* δ over mixed logit models including the controls from Study 1

Priming model with cue combination and decay



Results

Prime **Strength** ($\chi^2(1)=7.4$, $p < .01$)

Model fit by brute-force search and Quasi-Newton method. In the best model, **recent experience** is weighted **twice as strongly** as ***a-priori*** **expectations**.The best model also showed activation **decay** ($\delta=0.5$, cf. [19]).

Conclusions

(1) Syntactic priming is surprisal-sensitive [12,20]

(2) Syntactic priming may be an epiphenomenon of probabilistic maintenance, consistent with implicit learning accounts [6]

(3) Evidence is consistent with the assumption that language users are constantly updating their expectations based on recent experience (cf. [7]).

References

1. Anderson, J.R. and Lebiere, C. (1998). The atomic components of thought. Erlbaum.
2. Bannard, C. and Matthews, D. (2008). Stored Word Sequences in Language Learning: The Effect of Familiarity on Children's Repetition of Four-Word Combinations. Psychological Science. 19:3, 241-248.
3. Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. The Journal of the Acoustical Society of America. v.113.
4. Bernolet, S., Hartsuiker, R., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. Journal of Memory and Language.
5. Bock, K. (1986). Syntactic persistence in language production. Cognitive Psychology, 18(3), 355-387. New York, NY.
6. Bock, K., & Griffin, Z. (2000). The persistence of struc- tural priming: Transient activation or implicit learning. Journal of Experimental Psychology: General, 129(2), 177-192.
7. Clayards, M., Tanenhaus, M.K., Aslin, R.N., and Jacobs, R.A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. Cognition. 108:3. 804-809.
8. Ferreira, V. S. (2003). The persistence of optional complementizer mention: Why saying a "that" is not saying "that" at all. Journal of Memory and Language, 48, 379–398.
9. Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. Proceedings of NAACL. v.2, 159-166.
10. Huber, D. E., & O'Reilly, R. C. (2003). Persistence and accommodation in short-term priming and other perceptual paradigms: temporal segregati on through synaptic depression. Cognitive Science, 27(3), 403–430.
11. Huber, D. E., Shiffrin, R. M., Lyle, K. B., & Ruys, K. I. (2001). Perception and preference in short-term word priming. Psychological Review, 108, 149–182.
12. Jaeger, T. F., & Snider, N. (2008). Implicit learning and syntactic persistence: Surprisal and cumulativity. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society. pp.1061-1067. Austin, TX: Cognitive Science Society.
13. Kaschak, M., & Borreggine, K. (2007). Is long-term structural priming affected by patterns of experience with individual verbs? Journal of Memory and Language.
14. Kraljic, T., Brennan, S.E., and Samuel, A.G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. Cognition. 107:1, 54-81.
15. Levy, R. (2008). Expectation-based syntactic comprehension. Cognition. 106:3, 1126-1177.
16. Levy, R. and Jaeger, T.F. (2007). Speakers optimize information density through syntactic reduction. Advances in neural information processing systems 19. MIT.
17. MacDonald, M.C. (1999). Distributional information in language comprehension, production, and acquisition. Three puzzles and a moral. In B. MacWhinney (Ed.). The emergence of Language. 177-196. Mahwah, NJ: Erlbaum.
18. McDonald, S.A. and Shillcock, R.C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. Psychological Science. 14:6, 648-652. Blackwell.
19. Reitter, D., Moore, J., & Keller, F. (2006). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In Proceedings of the 28th annual conference of the cognitive science society.
20. Snider, N. (2008). An exemplar model of syntactic priming. Unpublished doctoral dissertation, Stanford University, Stanford, CA.