Modeling syntax acquisition via cognitively-constrained unsupervised grammar induction

Cory Shain[1], William Bryce[2], Lifeng Jin[1], Victoria Krakovna[3], Finale Doshi-Velez[3], Timothy Miller[4], William Schuler[1], and Lane Schwartz[2]

shain.3@osu.edu

[1]OSU,[2]UIUC,[3]Harvard,[4]Harvard Medical School & Boston Children's Hospital

Question

What can word distributions reveal about syntactic structure? How might human learners exploit this information?

Hypothesis

Unsupervised grammar induction (a machine-learning task) can discover latent syntax in word distributions and quantify (lower-bound) the learnability problem of natural language syntax.

Problem

Existing grammar induction techniques [13, 12, 1, 10] do not model (1) incremental left-corner parsing [5, 4, 7, 15] or (2) limited working memory [9, 2, 16]. They might exploit information unavailable to human learners.

References

[1] Christos Christodoulopoulos, Sharon Goldwater, and Mark Steedman. Turning the pipeline into a loop: Iterated unsupervised dependency parsing and PoS induction. In *NAACL-HLT Workshop on the Induction of Linguistic Structure*, pages 96–99, Montreal, Canada, 6 2012.

[2] Nelson Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24:87–185, 2001.

[3] Janet Dean Fodor and William Gregory Sakas. Evaluating models of parameter setting. In A. Brugos, L. Micciulla, and C. E. Smith, editors, *BUCLD 28*, Somerville, 2004. Cascadilla Press.

[4] Edward Gibson. *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. PhD thesis, Carnegie Mellon, 1991.

[5] Philip N. Johnson-Laird. *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press, Cambridge, MA, USA, 1983.

[6] Tom Kwiatkowski, Sharon Goldwater, Luke S. Zettlemoyer, and Mark Steedman. A probabilistic model of syntactic and semantic acquisition from child-directed utterances and their meanings. In *Proceedings of EACL 2012*, 2012.

[7] Richard L. Lewis and Shravan Vasishth. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3):375–419, 2005.

[8] Brian MacWhinney. *The CHILDES project: Tools for analyzing talk*. Lawrence Elrbaum Associates, Mahwah, NJ, third edition, 2000.

[9] George A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63:81–97, 1956.

[10] John Pate and Mark Johnson. Grammar induction from (lots of) words alone. In *COLING*, pages 23 – 32, 2016.

[11] Lisa Pearl and Jon Sprouse. Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition*, 20:23–68, 2013.

[12] Elias Ponvert, Jason Baldridge, and Katrin Erik. Simple unsupervised grammar induction from raw text with cascaded finite state models. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 1077–1086, Portland, Oregon, 6 2011.

[13] Yoav Seginer. Fast unsupervised incremental parsing. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 384–391, 2007.

[14] Cory Shain, William Bryce, Lifeng Jin, Victoria Krakovna, Finale Doshi-Velez, Timothy Miller, William Schuler, and Lane Schwartz. Memory-bounded left-corner unsupervised grammar induction on child-directed input. In *Proceedings of The 26th International Conference on Computational Linguistics*, pages 964–975, Osaka, 2016.

[15] Cory Shain, Marten van Schijndel, Richard Futrell, Edward Gibson, and William Schuler. Memory access during incremental sentence processing causes reading time latency. In *Proceedings of the Computational Linguistics for Linguistic Complexity Workshop*, pages 49–58. Association for Computational Linguistics, 2016.

[16] Julie A. Van Dyke and Clinton L. Johns. Memory interference as a determinant of language comprehension. *Language and Linguistics Compass*, 6(4):193–211, 2012.

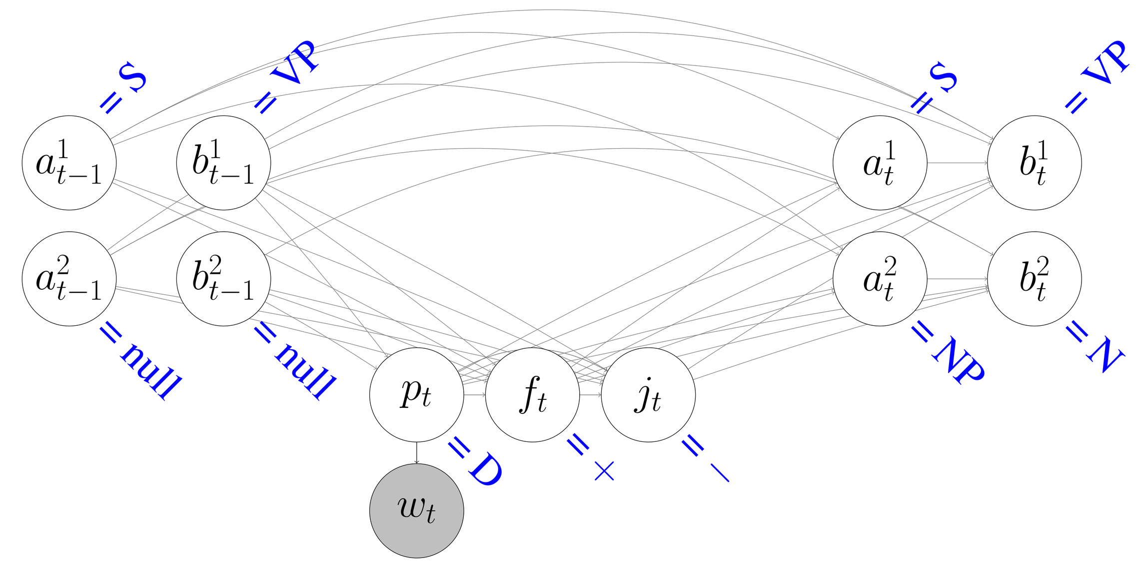
Our approach

Use a new memory-limited left-corner **unsupervised hierarchical hidden Markov model** (UHHMM) learner to discover English syntax from child-directed speech [14]. No universal grammar (cf. e.g. [3]) or semantic model (cf. e.g. [6]), which allows us to test cue utility of word distributions alone.

System design

Structure: Depth-limited left-corner unsupervised hierarchichal hidden Markov model
Training: Batch Gibbs sampling
Data: Child-directed English speech (Eve [8])
Parameters: $|A| = 4$, $|B| = 4$, $|P| = 8$, depth $= 2$
Evaluation standard: CHILDES Treebank [11]

Bayesian UHHMM left-corner parser



I gave the dog a bone.

a = label of active sign (being built)
b = label of awaited sign (needed to complete a)
p = part of speech
w = word (observed)
f = 'fork' decision (whether p completes b)
j = 'join' decision (whether bottom sign completes awaited sign above it).

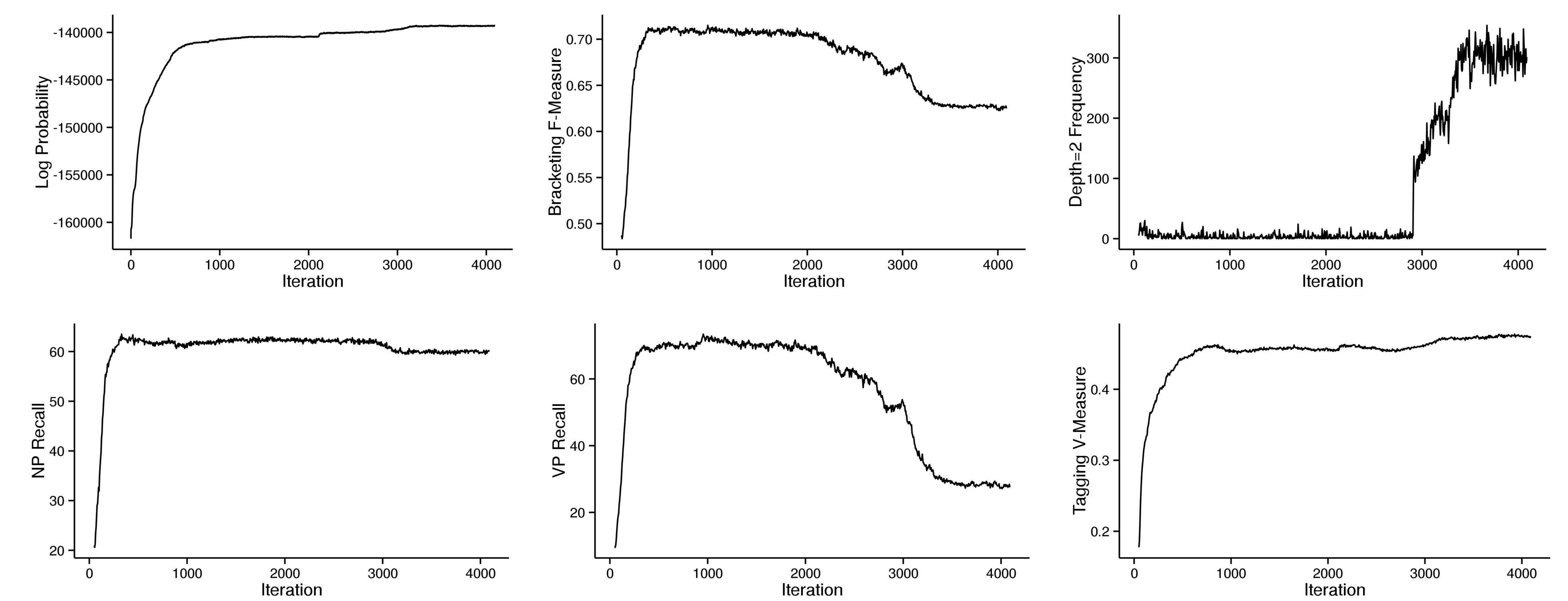
Gibbs sampling

Markov chain Monte Carlo sampling algorithm. Approximate inference of true posterior (cf. variational Bayes — exact inference of approximate posterior, e.g. [10]). For each sentence, computes posterior in a forward pass and samples parse in a backward pass.

Procedure:

```
randomly initialize probability models
iteration := 1
while iteration < maxIteration
    for sentence in data
        compute posterior (forward pass)
        sample from posterior (backward pass)
    endfor
    update models from sampled counts
    iteration += 1
endwhile
```

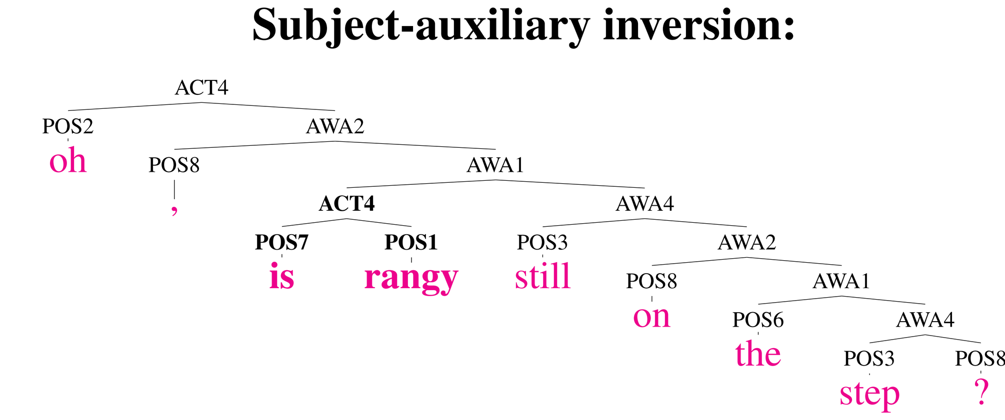
UHHMM learning curves



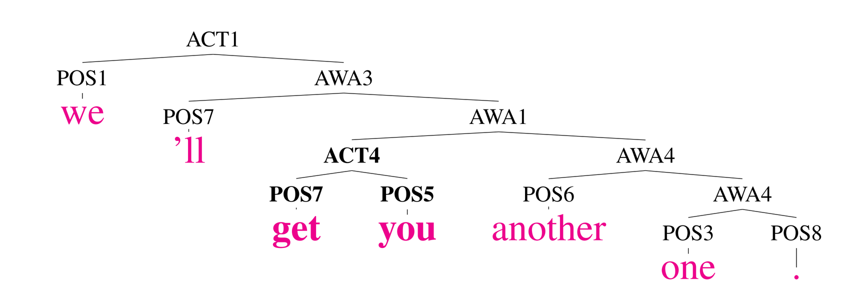
UHHMM makes rapid early progress, but loses accuracy as it starts to consider more complex parses. Noun phrases are easier to learn than verb phrases.

Actual parse examples

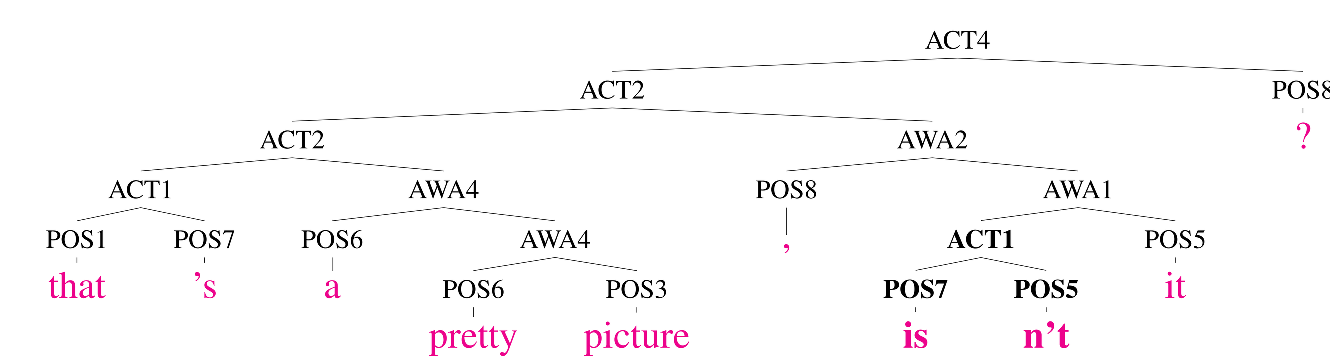
Subject-auxiliary inversion:



Ditransitive:



Contraction:



UHHMM uses depth 2 to learn linguistically-plausible constructions like those above. These are given flat representation in the gold trees, so our learner is not being rewarded for this insight.

Results

UHHMM ($F_1 = 62.47$) performs on par with BMMM+DMV[1] ($F_1 = 63.82$), a state-of-the-art competing system that does not model working memory limitations or incremental left-corner parsing. Learning curves reveal rapid early progress toward accurate parsing. UHHMM learns early to avoid center-embedding, then loses accuracy as it starts to seriously entertain center-embedded parses. The system has a much easier time learning noun phrases than verb phrases (verb phrases have more variable syntax). It also learns parts of speech, but not as well as a state-of-the-art unsupervised tagger (BMMM, $VM = 64.45$).

Conclusion

The system learns a lot of structure and makes linguistically interesting generalizations, but there is residue that may be difficult to learn without additional guidance (semantics, universal grammar, etc.). Future work on cognitively-constrained grammar induction may help resolve this question.

Acknowledgments: This project was sponsored by the Defense Advanced Research Projects Agency award #HR0011-15-2-0022. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.