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

Unsupervised grammar induction (inferring a grammar from raw text) is a machine learning task that strongly resembles the challenge faced by human infants of acquiring the syntax of their target language [12, 4, 24, 18]. It assumes no universal grammar (c.f. e.g. [8]) and therefore stands to shed light on the learnability of natural language syntax [3, 7] by quantifying the minimum utility of distributional statistics [22, 6] to syntax acquisition given particular assumptions about the input and the acquisition process. However, to our knowledge, no existing grammar induction system (e.g. [23, 20, 4]) models two likely features of human sentence processing: (1) constraints on working memory [16, 5, 15, 27] and (2) the use of a left-corner parsing strategy [11, 1, 10, 21, 26, 13, 25]. Furthermore, many grammar induction systems (e.g. [20, 4]) parse non-incrementally, potentially allowing them to benefit from information unavailable to humans during parsing. Grammar induction systems also tend to be evaluated on adult-directed newswire input, which likely differs in important ways from the input received by children. These considerations raise questions about the extent to which the syntactic generalizations learned by these systems are indeed learnable by humans, who face more severe constraints.

This work presents a new unsupervised grammar induction system that implements both limited working memory and a left-corner parsing strategy. Our system also attempts the more challenging learning problem of full joint acquisition of syntactic categorization and structural parsing ([23] and [20] only learn unlabeled bracketings, and [4] separates part of speech tagging and dependency parsing). The system updates its rule probabilities in order to estimate the most likely decision sequence by calculating and sampling from a posterior distribution at each time step [28, 24]. We simulate the syntax acquisition process in a left-corner learner by exposing the system to child-directed input, and evaluate the extent of its learning by comparing its parses to those of a human-annotated baseline [19]. To model infants' cognitive limitations [9] and small hypothesis space [17], we limit memory store depth to at most two disjoint derivation fragments and assume a small number of syntactic categories.[1] Our model lacks an explicit notion of world knowledge or reference, which allows us to probe the richness of distributional statistics alone as cues to syntactic structure during the acquisition and parsing process.

We evaluate this system on the child-directed sentences of the Eve section of the Brown corpus [2] distributed through the CHILDES database [14]. Our system achieves an accuracy (F1) score of 62.47, improving significantly ($p < 0.0001$) over a random baseline accuracy of 44.30.Competing systems without the aforementioned cognitive constraints ([23], [20], and [4]) performed near or below our level of accuracy, suggesting that cognitive constraints on parsing do not inhibit the syntax acquisition task, and in some cases may even aid it. In addition, our system uses center-embedding to discover linguistically interesting constructions, such as subject-auxiliary inversion (which has been argued to be too difficult to learn from data, see e.g. [3]). Our result shows that word distributions are informative to a cognitively-constrained learner. At the same time, there is much residual unlearned structure, some of which may be truly unlearnable, and some of which may be learnable under different assumptions about the input or the acquisition process. Further work on cognitively-constrained grammar induction may permit additional traction on this question.

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.

[1]We provide our left-corner system with four active categories (signs currently being built), four awaited categories (signs needed in order to complete a category), and eight parts of speech.

## References


[1] Steven P. Abney and Mark Johnson. Memory requirements and local ambiguities of parsing strategies. *J. Psycholinguistic Research*, 20(3):233–250, 1991.

[2] R. Brown. *A First Language*. Harvard University Press, Cambridge, MA, 1973.

[3] Noam Chomsky. *Language and Mind*. Harcourt, Brace & World, New York, 1968.

[4] 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.

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

[6] Jeffrey L. Elman. Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7:195–225, 1991.

[7] Janet Dean Fodor. Parsing to learn. *Journal of psycholinguistic research*, 27(3):339–374, 1998.

[8] 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.

[9] Susan E. Gathercole. The development of memory. *Journal of Child Psychology and Psychiatry*, 39:3–27, 1998.

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

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

[12] Dan Klein and Christopher D. Manning. Corpus-based induction of syntactic structure: Models of dependency and constituency. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, 2004.

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

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

[15] Brian McElree. Working memory and focal attention. *Journal of Experimental Psychology, Learning Memory and Cognition*, 27(3):817–835, 2001.

[16] 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.

[17] Elissa Newport. Maturational constraints on language learning. *Cognitive Science*, 14:11–28, 1990.

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

[19] 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.

[20] 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.

[21] Philip Resnik. Left-corner parsing and psychological plausibility. In *Proceedings of COLING*, pages 191–197, Nantes, France, 1992.

[22] Jenny R Saffran, Elizabeth K Johnson, Richard N Aslin, and Elissa L Newport. Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1):27–52, 1999.

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

[24] 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.

[25] 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.

[26] Edward Stabler. The finite connectivity of linguistic structure. In *Perspectives on Sentence Processing*, pages 303–336. Lawrence Erlbaum, 1994.

[27] 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.

[28] Jurgen Van Gael, Yunus Saatci, Yee Whye Teh, and Zoubin Ghahramani. Beam sampling for the infinite hidden Markov model. In *Proceedings of the 25th international conference on Machine learning*, pages 1088–1095. ACM, 2008.