

References

- Steven P Abney. 1987. *The English noun phrase in its sentential aspect*. Ph.D. thesis, Massachusetts Institute of Technology.
- P. Adriaans, M. Trautwein, and M. Vervoort. 2000. Towards high speed grammar induction on large text corpora. In *SOFSEM 2000, LNCS 1963*.
- J. K. Baker. 1979. Trainable grammars for speech recognition. In *Speech Communication Papers for the 97th Meeting of the Acoustical Society of America*.
- Taylor Berg-Kirkpatrick, Alexandre Bouchard-Côté, John DeNero, and Dan Klein. 2010. Painless unsupervised learning with features. In *NAACL*.
- Taylor Berg-Kirkpatrick and Dan Klein. 2010. Phylogenetic grammar induction. In *ACL*.
- Yonatan Bisk, Christos Christodoulopoulos, and Julia Hockenmaier. 2015. [Labeled grammar induction with minimal supervision](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 870–876, Beijing, China. Association for Computational Linguistics.
- Yonatan Bisk and Julia Hockenmaier. 2012. Simple robust grammar induction with combinatory categorial grammars. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, AAAI’12, page 1643–1649. AAAI Press.
- Yonatan Bisk and Julia Hockenmaier. 2013. [An HDP model for inducing Combinatory Categorical Grammars](#). *Transactions of the Association for Computational Linguistics*, 1:75–88.
- Phil Blunsom and Trevor Cohn. 2010. [Unsupervised induction of tree substitution grammars for dependency parsing](#). In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1204–1213, Cambridge, MA. Association for Computational Linguistics.
- Rens Bod. 2006a. [An all-subtrees approach to unsupervised parsing](#). In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 865–872, Sydney, Australia. Association for Computational Linguistics.
- Rens Bod. 2006b. [Unsupervised parsing with U-DOP](#). In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 85–92, New York City. Association for Computational Linguistics.
- Jiong Cai, Yong Jiang, and Kewei Tu. 2017. Crf autoencoder for unsupervised dependency parsing. In *the 2017 Conference on EMNLP*.
- Steven Cao, Nikita Kitaev, and Dan Klein. 2020. [Unsupervised parsing via constituency tests](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4798–4808, Online. Association for Computational Linguistics.
- Chaofan Chen, Oscar Li, Daniel Tao, Alina Barnett, Cynthia Rudin, and Jonathan K Su. 2019. This looks like that: deep learning for interpretable image recognition. In *Advances in neural information processing systems*, pages 8930–8941.
- Stanley F. Chen. 1995. [Bayesian grammar induction for language modeling](#). In *33rd Annual Meeting of the Association for Computational Linguistics*, pages 228–235, Cambridge, Massachusetts, USA. Association for Computational Linguistics.
- Jihun Choi, Kang Min Yoo, and Sang-goo Lee. 2018. Unsupervised learning of task-specific tree structures with tree-lstms. *AAAI*.
- A. Clark. 2007. Learning deterministic context free grammars: The omphalos competition. *Machine Learning*, 66.
- Alexander Clark. 2001. [Unsupervised induction of stochastic context-free grammars using distributional clustering](#). In *Proceedings of the ACL 2001 Workshop on Computational Natural Language Learning (ConLL)*.
- Shay B Cohen, Kevin Gimpel, and Noah A Smith. 2008. Logistic normal priors for unsupervised probabilistic grammar induction. In *Advances in Neural Information Processing Systems*, pages 321–328.
- Shay B Cohen and Noah A Smith. 2009. Shared logistic normal distributions for soft parameter tying in unsupervised grammar induction. In *NAACL*.
- Shay B Cohen and Noah A Smith. 2010. Covariance in unsupervised learning of probabilistic grammars. *The Journal of Machine Learning Research*.
- Trevor Cohn, Phil Blunsom, and Sharon Goldwater. 2010. [Inducing tree-substitution grammars](#). *Journal of Machine Learning Research*, 11(102):3053–3096.
- Michael Collins. 2003. [Head-driven statistical models for natural language parsing](#). *Computational Linguistics*, 29(4):589–637.
- Caio Corro and Ivan Titov. 2018. Differentiable perturb-and-parse: Semi-supervised parsing with a structured variational autoencoder. In *ICLR*.
- Hal Daumé III. 2009. Unsupervised search-based structured prediction. In *ICML*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL*.

- Thomas G Dietterich. 2000. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*. Springer.
- Andrew Drozdov, Patrick Verga, Mohit Yadav, Mohit Iyyer, and Andrew McCallum. 2019. Unsupervised latent tree induction with deep inside-outside recursive auto-encoders. In *NAACL*.
- Mengnan Du, Ninghao Liu, and Xia Hu. 2019. Techniques for interpretable machine learning. *Communications of the ACM*, 63(1):68–77.
- Xiangyu Duan, Min Zhang, and Wenliang Chen. 2013. Smoothing for bracketing induction. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, IJCAI '13*, page 2085–2091. AAAI Press.
- Long Duong, Trevor Cohn, Steven Bird, and Paul Cook. 2015. Cross-lingual transfer for unsupervised dependency parsing without parallel data. In *CoNLL*.
- Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A Smith. 2016. Recurrent neural network grammars. In *NAACL*.
- Douwe Gelling, Trevor Cohn, Phil Blunsom, and Joao Graça. 2012. The pascal challenge on grammar induction. In *the NAACL-HLT Workshop on the Induction of Linguistic Structure*.
- Jennifer Gillenwater, Kuzman Ganchev, Joao Graça, Fernando Pereira, and Ben Taskar. 2010. Sparsity in dependency grammar induction. In *ACL*.
- Jennifer Gillenwater, Kuzman Ganchev, Fernando Pereira, Ben Taskar, et al. 2011. Posterior sparsity in unsupervised dependency parsing. *Journal of Machine Learning Research*.
- E. Mark Gold. 1967. Language identification in the limit. *Information and Control*, 10(5):447–474.
- Yoav Goldberg and Omer Levy. 2014. word2vec explained: deriving mikolov et al.’s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*.
- Dave Golland, John DeNero, and Jakob Uszkoreit. 2012. A feature-rich constituent context model for grammar induction. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 17–22, Jeju Island, Korea. Association for Computational Linguistics.
- Edouard Grave and Noémie Elhadad. 2015. A convex and feature-rich discriminative approach to dependency grammar induction. In *ACL-IJCNLP*.
- Wenjuan Han, Yong Jiang, and Kewei Tu. 2017. Dependency grammar induction with neural lexicalization and big training data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1683–1688, Copenhagen, Denmark. Association for Computational Linguistics.
- Wenjuan Han, Yong Jiang, and Kewei Tu. 2019a. Enhancing unsupervised generative dependency parser with contextual information. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5315–5325, Florence, Italy. Association for Computational Linguistics.
- Wenjuan Han, Ge Wang, Yong Jiang, and Kewei Tu. 2019b. Multilingual grammar induction with continuous language identification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5728–5733, Hong Kong, China. Association for Computational Linguistics.
- Zellig S Harris. 1951. *Methods in structural linguistics*. University of Chicago Press.
- Zellig S Harris. 1954. Distributional structure. *Word*, 10(2-3):146–162.
- Junxian He, Graham Neubig, and Taylor Berg-Kirkpatrick. 2018. Unsupervised learning of syntactic structure with invertible neural projections. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1292–1302, Brussels, Belgium. Association for Computational Linguistics.
- Junxian He, Zhisong Zhang, Taylor Berg-Kirkpatrick, and Graham Neubig. 2019. Cross-lingual syntactic transfer through unsupervised adaptation of invertible projections. In *ACL*.
- William P Headden III, Mark Johnson, and David McClosky. 2009. Improving unsupervised dependency parsing with richer contexts and smoothing. In *NAACL*.
- Tomoharu Iwata, Daichi Mochihashi, and Hiroshi Sawada. 2010. Learning common grammar from multilingual corpus. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 184–188. Association for Computational Linguistics.
- Eric Jang, Shixiang Gu, and Ben Poole. 2017. Categorical reparametrization with gumble-softmax. In *ICLR*.
- Zixia Jia, Youmi Ma, Jiong Cai, and Kewei Tu. 2020. Semi-supervised semantic dependency parsing using CRF autoencoders. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6795–6805, Online. Association for Computational Linguistics.
- Yong Jiang, Wenjuan Han, and Kewei Tu. 2016. Unsupervised neural dependency parsing. In *EMNLP*.
- Yong Jiang, Wenjuan Han, and Kewei Tu. 2017. Combining generative and discriminative approaches to unsupervised dependency parsing via dual decomposition. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1689–1694, Copenhagen, Denmark. Association for Computational Linguistics.

- Yong Jiang, Wenjuan Han, and Kewei Tu. 2019. [A regularization-based framework for bilingual grammar induction](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1423–1428, Hong Kong, China. Association for Computational Linguistics.
- Lifeng Jin, Finale Doshi-Velez, Timothy Miller, William Schuler, and Lane Schwartz. 2018a. [Depth-bounding is effective: Improvements and evaluation of unsupervised PCFG induction](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2721–2731, Brussels, Belgium. Association for Computational Linguistics.
- Lifeng Jin, Finale Doshi-Velez, Timothy Miller, William Schuler, and Lane Schwartz. 2018b. [Unsupervised grammar induction with depth-bounded PCFG](#). *Transactions of the Association for Computational Linguistics*, 6:211–224.
- Lifeng Jin, Finale Doshi-Velez, Timothy Miller, Lane Schwartz, and William Schuler. 2019. [Unsupervised learning of PCFGs with normalizing flow](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2442–2452, Florence, Italy. Association for Computational Linguistics.
- Lifeng Jin and William Schuler. 2020. [Grounded PCFG induction with images](#). In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 396–408, Suzhou, China. Association for Computational Linguistics.
- Mark Johnson, Thomas Griffiths, and Sharon Goldwater. 2007. [Bayesian inference for PCFGs via Markov chain Monte Carlo](#). In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 139–146, Rochester, New York. Association for Computational Linguistics.
- Taeuk Kim, Jihun Choi, Daniel Edmiston, and Sangwoo Lee. 2020. [Are pre-trained language models aware of phrases? simple but strong baselines for grammar induction](#). In *International Conference on Learning Representations*.
- Yoon Kim, Carl Denton, Luong Hoang, and Alexander M Rush. 2017. Structured attention networks. *ICLR*.
- Yoon Kim, Chris Dyer, and Alexander Rush. 2019a. [Compound probabilistic context-free grammars for grammar induction](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2369–2385, Florence, Italy. Association for Computational Linguistics.
- Yoon Kim, Alexander M Rush, Lei Yu, Adhiguna Kuncoro, Chris Dyer, and Gábor Melis. 2019b. Unsupervised recurrent neural network grammars. In *NAACL*.
- Dan Klein and Christopher D. Manning. 2002. [A generative constituent-context model for improved grammar induction](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 128–135, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.
- Dan Klein and Christopher D. Manning. 2004. Corpus-based induction of syntactic structure: Models of dependency and constituency. In *ACL*.
- Adhiguna Kuncoro, Lingpeng Kong, Daniel Fried, Dani Yogatama, Laura Rimell, Chris Dyer, and Phil Blunsom. 2020. [Syntactic structure distillation pre-training for bidirectional encoders](#). *Transactions of the Association for Computational Linguistics*, 8(0):776–794.
- Kenichi Kurihara and Taisuke Sato. 2006. Variational Bayesian grammar induction for natural language. In *ICGI 2006*, volume 4201 of *LNAI*, pages 84–96.
- Pat Langley and Sean Stromsten. 2000. Learning context-free grammars with a simplicity bias. In *ECML*, pages 220–228.
- Phong Le and Willem Zuidema. 2015. Unsupervised dependency parsing: Let’s use supervised parsers. In *NAACL*.
- Bowen Li, Jianpeng Cheng, Yang Liu, and Frank Keller. 2019. Dependency grammar induction with a neural variational transition-based parser. In *AAAI*.
- Bowen Li, Taeuk Kim, Reinald Kim Amplayo, and Frank Keller. 2020a. [Heads-up! unsupervised constituency parsing via self-attention heads](#). In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 409–424, Suzhou, China. Association for Computational Linguistics.
- Jun Li, Yifan Cao, Jiong Cai, Yong Jiang, and Kewei Tu. 2020b. [An empirical comparison of unsupervised constituency parsing methods](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3278–3283, Online. Association for Computational Linguistics.
- Zhao Li and Kewei Tu. 2020. [Unsupervised cross-lingual adaptation of dependency parsers using CRF autoencoders](#). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2127–2133, Online. Association for Computational Linguistics.
- Kai Liu, Yajuan Lü, Wenbin Jiang, and Qun Liu. 2013. Bilingually-guided monolingual dependency grammar induction. In *ACL*.

- Xuezhe Ma and Fei Xia. 2014. Unsupervised dependency parsing with transferring distribution via parallel guidance and entropy regularization. In *ACL*.
- Christopher D. Manning. 2017. Last words: Computational linguistics and deep learning.
- Diego Marcheggiani and Ivan Titov. 2017. [Encoding sentences with graph convolutional networks for semantic role labeling](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1506–1515, Copenhagen, Denmark. Association for Computational Linguistics.
- David Mareček and Milan Straka. 2013. [Stop-probability estimates computed on a large corpus improve unsupervised dependency parsing](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 281–290, Sofia, Bulgaria. Association for Computational Linguistics.
- David Mareček and Zdeněk Žabokrtský. 2012. [Exploiting reducibility in unsupervised dependency parsing](#). In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 297–307, Jeju Island, Korea. Association for Computational Linguistics.
- André F. T. Martins, Tsvetomila Mihaylova, Nikita Nangia, and Vlad Niculae. 2019. [Latent structure models for natural language processing](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 1–5, Florence, Italy. Association for Computational Linguistics.
- Ryan McDonald, Fernando Pereira, Kiril Ribarov, and Jan Hajič. 2005. Non-projective dependency parsing using spanning tree algorithms. In *EMNLP*.
- Ryan McDonald, Slav Petrov, and Keith Hall. 2011. Multi-source transfer of delexicalized dependency parsers. In *EMNLP*.
- Alexander Mott, Daniel Zoran, Mike Chrzanowski, Daan Wierstra, and Danilo Jimenez Rezende. 2019. Towards interpretable reinforcement learning using attention augmented agents. In *Advances in Neural Information Processing Systems*, pages 12350–12359.
- Noriki Nishida and Hideki Nakayama. 2020. Unsupervised discourse constituency parsing using Viterbi EM. *TACL*, 8:215–230.
- Hiroshi Noji, Yusuke Miyao, and Mark Johnson. 2016. Using left-corner parsing to encode universal structural constraints in grammar induction. In *EMNLP*.
- John K Pate and Mark Johnson. 2016. [Grammar induction from \(lots of\) words alone](#). In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 23–32, Osaka, Japan. The COLING 2016 Organizing Committee.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *EMNLP (EMNLP)*.
- Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proc. of NAACL*.
- Hoifung Poon and Pedro Domingos. 2011. Sum-product networks: A new deep architecture. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence, UAI’11*, page 337–346, Arlington, Virginia, USA. AUAI Press.
- Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8).
- Michael Roth and Mirella Lapata. 2016. [Neural semantic role labeling with dependency path embeddings](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1192–1202, Berlin, Germany. Association for Computational Linguistics.
- Alexander M Rush and MJ Collins. 2012. A tutorial on dual decomposition and lagrangian relaxation for inference in natural language processing. *Journal of Artificial Intelligence Research*, 45.
- Roy Schwartz, Omri Abend, Roi Reichart, and Ari Rapoport. 2011. [Neutralizing linguistically problematic annotations in unsupervised dependency parsing evaluation](#). In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 663–672, Portland, Oregon, USA. Association for Computational Linguistics.
- James Scicluna and Colin de la Higuera. 2014. [PCFG induction for unsupervised parsing and language modelling](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1353–1362, Doha, Qatar. Association for Computational Linguistics.
- Yoav Seginer. 2007. [Fast unsupervised incremental parsing](#). In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 384–391, Prague, Czech Republic. Association for Computational Linguistics.
- Yikang Shen, Zhouhan Lin, Chin wei Huang, and Aaron Courville. 2018. [Neural language modeling by jointly learning syntax and lexicon](#). In *International Conference on Learning Representations*.
- Yikang Shen, Shawn Tan, Alessandro Sordoni, and Aaron Courville. 2019. [Ordered neurons: Integrating tree structures into recurrent neural networks](#). In *International Conference on Learning Representations*.

- Haoyue Shi, Karen Livescu, and Kevin Gimpel. 2020. [On the role of supervision in unsupervised constituency parsing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7611–7621, Online. Association for Computational Linguistics.
- Haoyue Shi, Jiayuan Mao, Kevin Gimpel, and Karen Livescu. 2019. [Visually grounded neural syntax acquisition](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1842–1861, Florence, Italy. Association for Computational Linguistics.
- Xing Shi, Inkit Padhi, and Kevin Knight. 2016. [Does string-based neural MT learn source syntax?](#) In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1526–1534, Austin, Texas. Association for Computational Linguistics.
- Noah A. Smith and Jason Eisner. 2004. [Annealing techniques for unsupervised statistical language learning](#). In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, pages 486–493, Barcelona, Spain.
- Noah A. Smith and Jason Eisner. 2006. [Annealing structural bias in multilingual weighted grammar induction](#). In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 569–576, Sydney, Australia. Association for Computational Linguistics.
- Benjamin Snyder, Tahira Naseem, and Regina Barzilay. 2009. [Unsupervised multilingual grammar induction](#). In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 73–81, Suntec, Singapore. Association for Computational Linguistics.
- Anders Søgaard. 2012. Unsupervised dependency parsing without training. *Natural Language Engineering*, 18(2):187.
- Zach Solan, David Horn, Eytan Ruppín, and Shimon Edelman. 2005. Unsupervised learning of natural languages. *Proc. Natl. Acad. Sci.*, 102(33):11629–11634.
- Valentin I. Spitzkovsky, Hiyan Alshawi, Angel X. Chang, and Daniel Jurafsky. 2011a. [Unsupervised dependency parsing without gold part-of-speech tags](#). In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1281–1290, Edinburgh, Scotland, UK. Association for Computational Linguistics.
- Valentin I Spitzkovsky, Hiyan Alshawi, and Daniel Jurafsky. 2011b. Lateen em: Unsupervised training with multiple objectives, applied to dependency grammar induction. In *EMNLP*.
- Valentin I. Spitzkovsky, Hiyan Alshawi, and Daniel Jurafsky. 2011c. [Punctuation: Making a point in unsupervised dependency parsing](#). In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning*, pages 19–28, Portland, Oregon, USA. Association for Computational Linguistics.
- Valentin I Spitzkovsky, Hiyan Alshawi, Daniel Jurafsky, and Christopher D Manning. 2010. Viterbi training improves unsupervised dependency parsing. In *CoNLL*.
- Andreas Stolcke and Stephen M. Omohundro. 1994. Inducing probabilistic grammars by Bayesian model merging. In *ICGI*, pages 106–118.
- Emma Strubell, Patrick Verga, Daniel Andor, David Weiss, and Andrew McCallum. 2018. [Linguistically-informed self-attention for semantic role labeling](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5027–5038, Brussels, Belgium. Association for Computational Linguistics.
- Ian Tenney, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R Thomas McCoy, Najoung Kim, Benjamin Van Durme, Sam Bowman, Dipanjan Das, and Ellie Pavlick. 2019. [What do you learn from context? probing for sentence structure in contextualized word representations](#). In *International Conference on Learning Representations*.
- Kewei Tu and Vasant Honavar. 2008. Unsupervised learning of probabilistic context-free grammar using iterative biclustering. In *Proceedings of 9th International Colloquium on Grammatical Inference (ICGI 2008)*, LNCS 5278.
- Kewei Tu and Vasant Honavar. 2012. Unambiguity regularization for unsupervised learning of probabilistic grammars. In *EMNLP-CoNLL*.
- Kewei Tu, Maria Pavlovskaya, and Song-Chun Zhu. 2013. [Unsupervised structure learning of stochastic and-or grammars](#). In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.
- Ge Wang and Kewei Tu. 2020. [Semi-supervised dependency parsing with arc-factored variational autoencoding](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2485–2496, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Zhiyong Wu, Yun Chen, Ben Kao, and Qun Liu. 2020. [Perturbed masking: Parameter-free probing for analyzing and interpreting BERT](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4166–4176, Online. Association for Computational Linguistics.
- Dani Yogatama, Phil Blunsom, Chris Dyer, Edward Grefenstette, and Wang Ling. 2016. Learning to compose words into sentences with reinforcement learning. In *ICLR*.

- Juntao Yu, Mohab El-karef, and Bernd Bohnet. 2015. Domain adaptation for dependency parsing via self-training. In *the 14th International Conference on Parsing Technologies*.
- Menno van Zaanen. 2000. [ABL: Alignment-based learning](#). In *COLING 2000 Volume 2: The 18th International Conference on Computational Linguistics*.
- Quanshi Zhang, Ruiming Cao, Feng Shi, Ying Nian Wu, and Song-Chun Zhu. 2018. Interpreting cnn knowledge via an explanatory graph. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- Xiao Zhang and Dan Goldwasser. 2020. [Semi-supervised parsing with a variational autoencoding parser](#). In *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*, pages 40–47, Online. Association for Computational Linguistics.
- Yanpeng Zhao and Ivan Titov. 2020. [Visually grounded compound PCFGs](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4369–4379, Online. Association for Computational Linguistics.
- Hao Zhu, Yonatan Bisk, and Graham Neubig. 2020. [The return of lexical dependencies: Neural lexicalized pcfgs](#). *Transactions of the Association for Computational Linguistics*, 8:647–661.