Recent Advances in Dependency Parsing

Tutorial, EACL, April 27th, 2014

Ryan McDonald[1] Joakim Nivre[2]

[1]Google Inc., USA/UK
E-mail: ryanmcd@google.com

[2]Uppsala University, Sweden
E-mail: joakim.nivre@lingfil.uu.se

Overview of the Tutorial

- Introduction to Dependency Parsing (Joakim)
- Graph-based parsing post-2008 (Ryan)
- Transition-based parsing post-2008 (Joakim)
- **Summary and final thoughts** (Ryan)

Topics Not Discussed

Unsupervised Learning

Learn models from unlabeled data only

- Dependency Model with Valency (DMV) [Klein and Manning 2004]
  - Plus extensions [Cohen et al. 2008, Headden III et al. 2009]
- Adding in prior knowledge/constraints
  - Sparsity over head-modifier POS combinations [Gillenwater et al. 2010]
  - Universal linguistic knowledge [Naseem et al. 2010]
  - Multi-lingual induction [Berg-Kirkpatrick and Klein 2010, Cohen and Smith 2009]

Semi-Supervised Learning

Learn models from labeled *and* unlabeled data

- Co-training [Sagae and Tsujii 2007]
- Tri-training [Søgaard and Rishøj 2010a]
- Up/self-training [Petrov et al. 2010]
- Cluster features [Koo et al. 2008, Täckström et al. 2012]
- Structured conditional model [Suzuki et al. 2009]
- Web-derived features [Bansal and Klein 2011]
- Dependency language model [Chen et al. 2012]
- Meta-features [Chen et al. 2013]

Cross-Lingual Learning

Learn models from foreign or parallel language resources

- Projection [Hwa et al. 2005, Ganchev et al. 2009]
- Delexicalized transfer [Zeman and Resnik 2008]
- Multi-source training [McDonald et al. 2011, Cohen et al. 2011]
- Linguistic priors [Naseem et al. 2012, Täckström et al. 2013]
- Cross-lingual resources [Täckström et al. 2012, Durrett et al. 2012]

- Significantly higher accuracies than unsupervised learning [McDonald et al. 2011]

Learning with Approximate Search

- Learning algorithms assume exact search
- Pruning, beam search and other approximations break this
- Huang et al. [2012]: principled method for perceptron with approximate search
- Transition-based parsing [Huang et al. 2012]
  - English UAS: 92.1 $\rightarrow$ 92.2
  - Speeds up training by factor of 3
- Graph-based parsing [Zhang et al. 2013]
  - English LAS: 92.92 $\rightarrow$ 93.64
  - English LAS: 90.35 $\rightarrow$ 91.28
  - Bottom-up parsing has larger search space

Domain Adaptation

- Dependency parsers are subject to domain shift
  - WSJ LAS → QTB LAS: 86.4 → 67.0 [Petrov et al. 2010]
  - WSJ LAS → Web LAS: 91.5 → 83.4 [Petrov and McDonald 2012]
- Ensembles and self-training [Sagae and Tsujii 2007]
- Datapoint selection [Kawahara and Uchimoto 2008]
- Grammars + statistical parsers

  [Zhang and Wang 2009, Petrov et al. 2010]
- Tri-training [Søgaard and Rishøj 2010b]
- Training with domain specific loss functions [Hall et al. 2011]
- Shared tasks

  [Nivre et al. 2007, Dell'Orletta et al. 2011, Petrov and McDonald 2012]

Parsing General Graphs

- Tree constraint often just a computational convenience
- Conceptually, dependency graphs don't need to be trees
- Directed arcs between words can encode:
  - Raising and control structures
  - More direct co-ordination structures
  - Traces and wh-movement
  - ...
- Such structures are more semantic in nature

- McDonald and Pereira [2006] parse DAGs with approximate graph-based inference
- Sagae and Tsujii [2008] extend transition-based system to parse DAGs
- Many other solutions in CoNLL 2008 and 2009 shared tasks

  [Surdeanu et al. 2008, Hajič et al. 2009]

Using Phrase-Structure Parsers

- Lexicalized phrase-structure parsers produce dependencies [Collins 1999, Charniak 2000]
- Phrase-structure parsers can also produce dependencies via post-processing [Cer et al. 2010]
  - Latter is more accurate [McDonald et al. 2005]
  - Bias in post-processing [Petrov and McDonald 2012]
- Dependency to phrase-structure treebank conversion [Collins et al. 1999]
- Coppola and Steedman [2013]: cube-pruned phrase-structure parser with dependency features
  - Highest reported En scores for both phrase-structure and dependency evalualtions
  - Caveat: conversion heuristics to generate dependencies
  - Can combine with cube-pruned dependency parsers

Improved Evaluations

- LAS/UAS evaluations are useful, but
  - Treat all errors as equal
  - Don't say anything about downstream performance
  - Only allow comparisons on single annotation scheme
- Targeted dependency evaluations
  - Long-distance and/or implicit dependencies [Rimell et al. 2009, Nivre et al. 2010]
- Downstream task evals [Miyao et al. 2008]
  - Training to optimize task specific evals [Hall et al. 2011]
- Annotation-scheme independent evaluations [Tsarfaty et al. 2011]

Final Thoughts

Where do we stand?

LAS: 83.8 v. 83.6
[McDonald & Nivre 2007]

2008

Graph-based Parsers	Transition-based Parsers
Global Inference	Local Inference
Global Learning	Local Learnng
Local Feature Scope	Global Feature Scope

Graph-based	Transition-based
higher-order chart parsing	beam search
pruning	perceptron
ILP	dynamic oracles
dual decomp	dynamic programming
mildly non-projective	more features
etc.	etc.

2014

Graph-based Parsers	Transition-based Parsers
Global Inference	Global Inference
Global Learnng	Global Learnng
Global Feature Scope	Global Feature Scope

LAS: 85.8 v. 85.5
[Zhang et al. 2013]

Evaluated on overlapping 9 languages in studies

Where do we stand?

- 2008
  - Graph-based and transition-based have near identical accuracies [Buchholz and Marsi 2006]
  - But very different errors [McDonald and Nivre 2007]
  - And errors can be correlated with model properties
- 2014
  - 2008+: Attempts to address model short-comings
  - Models have converged (structured prediction, rich features, heuristic inference)
  - Accuracies again nearly identical [Zhang et al. 2013]

Where do we stand?

- 2008
  - Graph-based and transition-based have near identical accuracies [Buchholz and Marsi 2006]
  - But very different errors [McDonald and Nivre 2007]
  - And errors can be correlated with model properties
- 2014
  - 2008+: Attempts to address model short-comings
  - Models have converged (structured prediction, rich features, heuristic inference)
  - Accuracies again nearly identical [Zhang et al. 2013]

Do graph-based and transition-based parsers still make different kinds of errors?

Where do we stand?

- 2008
  - Graph-based and transition-based have near identical accuracies [Buchholz and Marsi 2006]
  - But very different errors [McDonald and Nivre 2007]
  - And errors can be correlated with model properties
- 2014
  - 2008+: Attempts to address model short-comings
  - Models have converged (structured prediction, rich features, heuristic inference)
  - Accuracies again nearly identical [Zhang et al. 2013]

Are the remaining model differences meaningful?
Formal/statistical power? Empirically?

References and Further Reading

- Mohit Bansal and Dan Klein. 2011. Web-scale features for full-scale parsing. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 693–702. Association for Computational Linguistics.
- Taylor Berg-Kirkpatrick and Dan Klein. 2010. Phylogenetic grammar induction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1288–1297. Association for Computational Linguistics.
- Sabine Buchholz and Erwin Marsi. 2006. CoNLL-X shared task on multilingual dependency parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, pages 149–164.
- Daniel M Cer, Marie-Catherine De Marneffe, Daniel Jurafsky, and Christopher D Manning. 2010. Parsing to stanford dependencies: Trade-offs between speed and accuracy. In *LREC*.
- Eugene Charniak. 2000. A maximum-entropy-inspired parser. In *Proceedings of the First Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 132–139.
- Wenliang Chen, Min Zhang, and Haizhou Li. 2012. Utilizing dependency language models for graph-based dependency parsing models. In *Proceedings of the 50th*

Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1, pages 213–222. Association for Computational Linguistics.

- Wenliang Chen, Min Zhang, and Yue Zhang. 2013. Semi-supervised feature transformation for dependency parsing. In *Proc. EMNLP*.
- Shay B Cohen and Noah A Smith. 2009. Shared logistic normal distributions for soft parameter tying in unsupervised grammar induction. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 74–82. Association for Computational Linguistics.
- 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, Dipanjan Das, and Noah A Smith. 2011. Unsupervised structure prediction with non-parallel multilingual guidance. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 50–61. Association for Computational Linguistics.
- Michael Collins, Jan Hajič, Lance Ramshaw, and Christoph Tillmann. 1999. A statistical parser for Czech. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 505–512.

- Michael Collins. 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Ph.D. thesis, University of Pennsylvania.
- Gregory F. Coppola and Mark Steedman. 2013. The effect of higher-order dependency features in discriminative phrase-structure parsing. In *Proc. ACL*.
- Felice Dell'Orletta, Simone Marchi, Simonetta Montemagni, Giulia Venturi, Tommaso Agnoloni, and Enrico Francesconi. 2011. Domain adaptation for dependency parsing at evalita 2011. In *Evaluation of Natural Language and Speech Tools for Italian*, pages 58–69. Springer.
- Greg Durrett, Adam Pauls, and Dan Klein. 2012. Syntactic transfer using a bilingual lexicon. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1–11. Association for Computational Linguistics.
- Kuzman Ganchev, Jennifer Gillenwater, and Ben Taskar. 2009. Dependency grammar induction via bitext projection constraints. 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: Volume 1-Volume 1*, pages 369–377. Association for Computational Linguistics.
- Jennifer Gillenwater, Kuzman Ganchev, João Graça, Fernando Pereira, and Ben Taskar. 2010. Sparsity in dependency grammar induction. In *Proceedings of the*

ACL 2010 Conference Short Papers, pages 194–199. Association for Computational Linguistics.

- Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, et al. 2009. The conll-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–18. Association for Computational Linguistics.
- Keith Hall, Ryan McDonald, Jason Katz-Brown, and Michael Ringgaard. 2011. Training dependency parsers by jointly optimizing multiple objectives. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1489–1499. Association for Computational Linguistics.
- William P Headden III, Mark Johnson, and David McClosky. 2009. Improving unsupervised dependency parsing with richer contexts and smoothing. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 101–109. Association for Computational Linguistics.
- Liang Huang, Suphan Fayong, and Yang Guo. 2012. Structured perceptron with inexact search. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 142–151.

- Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel texts. *Natural language engineering*, 11(3):311–325.
- Daisuke Kawahara and Kiyotaka Uchimoto. 2008. Learning reliability of parses for domain adaptation of dependency parsing. In *IJCNLP*, volume 8.
- Dan Klein and Christopher D. Manning. 2004. Corpus-based induction of syntactic structure: Models of dependency and constituency. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 479–486.
- Terry Koo, Xavier Carreras, and Michael Collins. 2008. Simple semi-supervised dependency parsing. In *Proc. ACL*.
- Ryan McDonald and Joakim Nivre. 2007. Characterizing the errors of data-driven dependency parsing models. In *Proceedings of the Join Conference on Empirical Methods in Natural Language Processing and the Conference on Computational Natural Language Learning (EMNLP-CoNLL)*.
- Ryan McDonald and Fernando Pereira. 2006. Online learning of approximate dependency parsing algorithms. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 81–88.

- Ryan McDonald, Koby Crammer, and Fernando Pereira. 2005. Online large-margin training of dependency parsers. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 91–98.
- Ryan McDonald, Slav Petrov, and Keith Hall. 2011. Multi-source transfer of delexicalized dependency parsers. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 62–72. Association for Computational Linguistics.
- Yusuke Miyao, Rune Sætre, Kenji Sagae, Takuya Matsuzaki, and Jun'ichi Tsujii. 2008. Task-oriented evaluation of syntactic parsers and their representations. In *ACL*, volume 8, pages 46–54.
- Tahira Naseem, Harr Chen, Regina Barzilay, and Mark Johnson. 2010. Using universal linguistic knowledge to guide grammar induction. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1234–1244. Association for Computational Linguistics.
- Tahira Naseem, Regina Barzilay, and Amir Globerson. 2012. Selective sharing for multilingual dependency parsing. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 629–637. Association for Computational Linguistics.

- J Nivre, J Hall, S Kübler, R McDonald, J Nilsson, S Riedel, and D Yuret. 2007. The conll 2007 shared task on dependency parsing. In *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL*, pages 915–932. sn.
- Joakim Nivre, Laura Rimell, Ryan McDonald, and Carlos Gómez-Rodríguez. 2010. Evaluation of dependency parsers on unbounded dependencies. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 833–841. Association for Computational Linguistics.
- Slav Petrov and Ryan McDonald. 2012. Overview of the 2012 shared task on parsing the web. In *Notes of the First Workshop on Syntactic Analysis of Non-Canonical Language (SANCL)*, volume 59.
- Slav Petrov, Pi-Chuan Chang, Michael Ringgaard, and Hiyan Alshawi. 2010. Uptraining for accurate deterministic question parsing. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 705–713. Association for Computational Linguistics.
- Laura Rimell, Stephen Clark, and Mark Steedman. 2009. Unbounded dependency recovery for parser evaluation. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, pages 813–821. Association for Computational Linguistics.

- Kenji Sagae and Jun'ichi Tsujii. 2007. Dependency parsing and domain adaptation with lr models and parser ensembles. In *EMNLP-CoNLL*, volume 2007, pages 1044–1050.
- Kenji Sagae and Jun'ichi Tsujii. 2008. Shift-reduce dependency dag parsing. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 753–760. Association for Computational Linguistics.
- Anders Søgaard and Christian Rishøj. 2010a. Semi-supervised dependency parsing using generalized tri-training. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 1065–1073. Association for Computational Linguistics.
- Anders Søgaard and Christian Rishøj. 2010b. Semi-supervised dependency parsing using generalized tri-training. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 1065–1073. Association for Computational Linguistics.
- Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. 2008. The conll-2008 shared task on joint parsing of syntactic and semantic dependencies. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pages 159–177. Association for Computational Linguistics.
- Jun Suzuki, Hideki Isozaki, Xavier Carreras, and Michael Collins. 2009. An empirical study of semi-supervised structured conditional models for dependency parsing. In *Proceedings of the 2009 Conference on Empirical Methods in Natural

Language Processing: Volume 2-Volume 2, pages 551–560. Association for Computational Linguistics.

- Oscar Täckström, Ryan McDonald, and Jakob Uszkoreit. 2012. Cross-lingual word clusters for direct transfer of linguistic structure. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 477–487. Association for Computational Linguistics.
- Oscar Täckström, Ryan McDonald, and Joakim Nivre. 2013. Target language adaptation of discriminative transfer parsers. In *Proceedings of NAACL*.
- Reut Tsarfaty, Joakim Nivre, and Evelina Ndersson. 2011. Evaluating dependency parsing: robust and heuristics-free cross-annotation evaluation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 385–396. Association for Computational Linguistics.
- Daniel Zeman and Philip Resnik. 2008. Cross-language parser adaptation between related languages. In *IJCNLP*, pages 35–42.
- Yi Zhang and Rui Wang. 2009. Cross-domain dependency parsing using a deep linguistic grammar. 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: Volume 1-Volume 1*, pages 378–386. Association for Computational Linguistics.

- Liang Zhang, Huang, Kai Zhao, and Ryan McDonald. 2013. Online learning for inexact hypergraph search. In *Proceedings of Empirical Methods in Natural Language Processing*.