

On the Role of Annotation in Data-Driven Dependency Parsing

Joakim Nivre

Uppsala University, Sweden

Introduction

- ▶ Supervised dependency parsers feed on annotation
- ▶ How does annotation influence parsing performance?
- ▶ Formal properties:
 - ▶ Expressivity and complexity of representations
 - ▶ Impact on parsing efficiency
- ▶ Substantive properties:
 - ▶ Theoretical linguistic assumptions
 - ▶ Impact on parsing accuracy

Formal Properties

- ▶ What class of formal objects are suitable for dependency annotation and parsing?
- ▶ Some candidates:
 - ▶ Dependency trees
 - ▶ Dependency forests
 - ▶ Directed acyclic dependency graphs
 - ▶ Arbitrary dependency graphs
- ▶ Mono- or multistratal annotation/parsing?
- ▶ Null elements?

Dependency Trees and Projectivity

- ▶ Projective dependency trees:
 - ▶ Exact inference in polynomial time
 - ▶ Deterministic parsing in $O(n)$ worst-case time
- ▶ Arbitrary dependency trees:
 - ▶ Exact inference intractable (except for arc-factored models)
 - ▶ Deterministic parsing in $O(n^2)$ worst-case time
- ▶ Practical running time depends on several factors

Case Study: Transition-Based Parsing

- ▶ Deterministic transition-based parsing [Nivre 2008]
- ▶ Parsing time determined by
 - ▶ complexity of transition system
 - ▶ classification time (cf. grammar constant)
- ▶ Experimental study:
 - ▶ CoNLL-X data (12 languages) [Buchholz and Marsi 2006]
 - ▶ SVM classifiers (all-versus-all)
 - ▶ Sentence length = n
 - ▶ Label set = L

Parser	Complexity	Time	LAS
Projective [Nivre 2003]			
Pseudo-projective [Nivre and Nilsson 2005]			
Non-projective [Nivre 2007]			

Case Study: Transition-Based Parsing

- ▶ Deterministic transition-based parsing [Nivre 2008]
- ▶ Parsing time determined by
 - ▶ complexity of transition system
 - ▶ classification time (cf. grammar constant)
- ▶ Experimental study:
 - ▶ CoNLL-X data (12 languages) [Buchholz and Marsi 2006]
 - ▶ SVM classifiers (all-versus-all)
 - ▶ Sentence length = n
 - ▶ Label set = L

Parser	Complexity	Time	LAS
Projective [Nivre 2003]	$O(n \cdot L ^2)$		
Pseudo-projective [Nivre and Nilsson 2005]	$O(n \cdot L ^4)$		
Non-projective [Nivre 2007]	$O(n^2 \cdot L ^2)$		

Case Study: Transition-Based Parsing

- ▶ Deterministic transition-based parsing [Nivre 2008]
- ▶ Parsing time determined by
 - ▶ complexity of transition system
 - ▶ classification time (cf. grammar constant)
- ▶ Experimental study:
 - ▶ CoNLL-X data (12 languages) [Buchholz and Marsi 2006]
 - ▶ SVM classifiers (all-versus-all)
 - ▶ Sentence length = n
 - ▶ Label set = L

Parser	Complexity	Time	LAS
Projective [Nivre 2003]	$O(n \cdot L ^2)$	1.00	
Pseudo-projective [Nivre and Nilsson 2005]	$O(n \cdot L ^4)$	1.59	
Non-projective [Nivre 2007]	$O(n^2 \cdot L ^2)$	1.63	

Case Study: Transition-Based Parsing

- ▶ Deterministic transition-based parsing [Nivre 2008]
- ▶ Parsing time determined by
 - ▶ complexity of transition system
 - ▶ classification time (cf. grammar constant)
- ▶ Experimental study:
 - ▶ CoNLL-X data (12 languages) [Buchholz and Marsi 2006]
 - ▶ SVM classifiers (all-versus-all)
 - ▶ Sentence length = n
 - ▶ Label set = L

Parser	Complexity	Time	LAS
Projective [Nivre 2003]	$O(n \cdot L ^2)$	1.00	79.3
Pseudo-projective [Nivre and Nilsson 2005]	$O(n \cdot L ^4)$	1.59	79.6
Non-projective [Nivre 2007]	$O(n^2 \cdot L ^2)$	1.63	79.8

Why Bother?

- ▶ Do we need non-projective dependencies?
- ▶ Two conversions of the Penn Treebank:
 - ▶ Old style = Non-local dependencies ignored
 - ▶ New style = Non-local dependencies recovered
- ▶ Experimental comparison [Johansson and Nugues 2007]
 - ▶ Parsing accuracy for MaltParser (LAS)
 - ▶ Classification accuracy for semantic roles (SRL)

	LAS	SRL
Old style		
New style		

- ▶ NB: New style used in CoNLL shared tasks 2007–2009

Why Bother?

- ▶ Do we need non-projective dependencies?
- ▶ Two conversions of the Penn Treebank:
 - ▶ Old style = Non-local dependencies ignored
 - ▶ New style = Non-local dependencies recovered
- ▶ Experimental comparison [Johansson and Nugues 2007]
 - ▶ Parsing accuracy for MaltParser (LAS)
 - ▶ Classification accuracy for semantic roles (SRL)

	LAS	SRL
Old style	90.3	
New style	87.6	

- ▶ NB: New style used in CoNLL shared tasks 2007–2009

Why Bother?

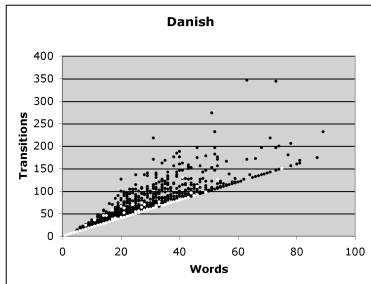
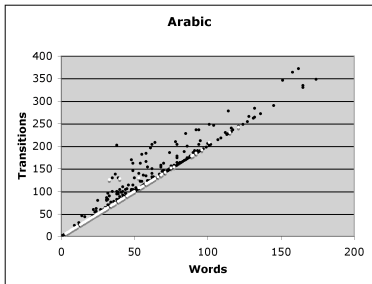
- ▶ Do we need non-projective dependencies?
- ▶ Two conversions of the Penn Treebank:
 - ▶ Old style = Non-local dependencies ignored
 - ▶ New style = Non-local dependencies recovered
- ▶ Experimental comparison [Johansson and Nugues 2007]
 - ▶ Parsing accuracy for MaltParser (LAS)
 - ▶ Classification accuracy for semantic roles (SRL)

	LAS	SRL
Old style	90.3	63.0
New style	87.6	72.5

- ▶ NB: New style used in CoNLL shared tasks 2007–2009

More Efficient Non-Projective Parsing

- ▶ Dependency parsing with online reordering [Nivre 2009]
 - ▶ Interleaved sorting and parsing
 - ▶ New transition for swapping input words
 - ▶ State-of-the-art results for non-projective dependency parsing
 - ▶ Expected linear time for representative inputs

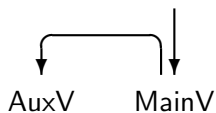
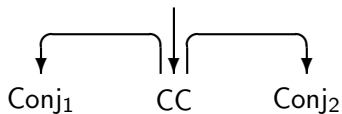


Substantive Properties

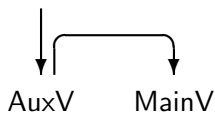
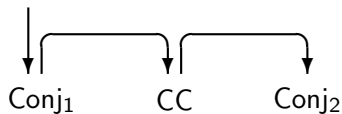
- ▶ How do we determine heads and dependents?
 - ▶ Conflicting criteria for many types of constructions
 - ▶ Even well-motivated annotations can be suboptimal for parsing
- ▶ Large body of work on PCFG transformations:
 - ▶ Parent annotation [Johnson 1998]
 - ▶ Head lexicalization [Collins 1999]
 - ▶ Horizontal and vertical markovization [Klein and Manning 2003]
 - ▶ State splitting [Petrov et al. 2006]
- ▶ Much less explored for dependency parsing

Case Study: Coordination and Verb Groups

Prague style



Mel'čuk style



Case Study: Coordination and Verb Groups

- ▶ Parsing experiments [Nilsson et al. 2006, Nilsson et al. 2007]:
 - ▶ Transform training data from Prague style to Mel'čuk style
 - ▶ Apply inverse transformation to parser output
- ▶ Data:
 - ▶ Prague Dependency Treebank [Hajič et al. 2001]
 - ▶ Slovene Dependency Treebank [Džeroski et al. 2006]
- ▶ Parser: MaltParser [Nivre et al. 2006a]

Transformation	Slovene	Czech
None		
Coordination		
Verb groups		

Case Study: Coordination and Verb Groups

- ▶ Parsing experiments [Nilsson et al. 2006, Nilsson et al. 2007]:
 - ▶ Transform training data from Prague style to Mel'čuk style
 - ▶ Apply inverse transformation to parser output
- ▶ Data:
 - ▶ Prague Dependency Treebank [Hajič et al. 2001]
 - ▶ Slovene Dependency Treebank [Džeroski et al. 2006]
- ▶ Parser: MaltParser [Nivre et al. 2006a]

Transformation	Slovene	Czech
None	77.3	83.4
Coordination		
Verb groups		

Case Study: Coordination and Verb Groups

- ▶ Parsing experiments [Nilsson et al. 2006, Nilsson et al. 2007]:
 - ▶ Transform training data from Prague style to Mel'čuk style
 - ▶ Apply inverse transformation to parser output
- ▶ Data:
 - ▶ Prague Dependency Treebank [Hajič et al. 2001]
 - ▶ Slovene Dependency Treebank [Džeroski et al. 2006]
- ▶ Parser: MaltParser [Nivre et al. 2006a]

Transformation	Slovene	Czech
None	77.3	83.4
Coordination	79.3 (+2.0)	85.5 (+2.1)
Verb groups		

Case Study: Coordination and Verb Groups

- ▶ Parsing experiments [Nilsson et al. 2006, Nilsson et al. 2007]:
 - ▶ Transform training data from Prague style to Mel'čuk style
 - ▶ Apply inverse transformation to parser output
- ▶ Data:
 - ▶ Prague Dependency Treebank [Hajič et al. 2001]
 - ▶ Slovene Dependency Treebank [Džeroski et al. 2006]
- ▶ Parser: MaltParser [Nivre et al. 2006a]

Transformation	Slovene	Czech
None	77.3	83.4
Coordination	79.3 (+2.0)	85.5 (+2.1)
Verb groups	77.9 (+0.6)	83.6 (+0.2)

Case Study: Coordination and Verb Groups

- ▶ Control experiment: Czech [Nilsson et al. 2007]
- ▶ Two parsers:
 - ▶ MaltParser [Nivre et al. 2006a]
 - ▶ MSTParser [McDonald et al. 2005]

Transformation	Malt	MST
None Coordination Verb groups		

Case Study: Coordination and Verb Groups

- ▶ Control experiment: Czech [Nilsson et al. 2007]
- ▶ Two parsers:
 - ▶ MaltParser [Nivre et al. 2006a]
 - ▶ MSTParser [McDonald et al. 2005]

Transformation	Malt	MST
None	83.4	84.5
Coordination		
Verb groups		

Case Study: Coordination and Verb Groups

- ▶ Control experiment: Czech [Nilsson et al. 2007]
- ▶ Two parsers:
 - ▶ MaltParser [Nivre et al. 2006a]
 - ▶ MSTParser [McDonald et al. 2005]

Transformation	Malt	MST
None	83.4	84.5
Coordination	85.5 (+2.1)	83.5 (-1.0)
Verb groups		

Case Study: Coordination and Verb Groups

- ▶ Control experiment: Czech [Nilsson et al. 2007]
- ▶ Two parsers:
 - ▶ MaltParser [Nivre et al. 2006a]
 - ▶ MSTParser [McDonald et al. 2005]

Transformation	Malt	MST
None	83.4	84.5
Coordination	85.5 (+2.1)	83.5 (-1.0)
Verb groups	83.6 (+0.2)	84.5 (± 0.0)

Case Study: Coordination and Verb Groups

- ▶ Control experiment: Czech [Nilsson et al. 2007]
- ▶ Two parsers:
 - ▶ MaltParser [Nivre et al. 2006a]
 - ▶ MSTParser [McDonald et al. 2005]

Transformation	Malt	MST
None	83.4	84.5
Coordination	85.5 (+2.1)	83.5 (-1.0)
Verb groups	83.6 (+0.2)	84.5 (± 0.0)

- ▶ No free lunch?

Optimizing Annotations

- ▶ All other things being equal, facilitate parsing
- ▶ Experiments on the Swedish Treebank [Nivre et al. 2006b]
- ▶ Minimizing dependency length helps:
 - ▶ Verb groups: AuxV \rightarrow MainV
 - ▶ Coordination: Conj₁ \rightarrow CC \rightarrow Conj₂
 - ▶ Prepositional phrases: Prep \rightarrow Noun
- ▶ Exception:
 - ▶ Subordinate clauses: Comp \leftarrow Verb
- ▶ Hypothesis:
 - ▶ Consistency more important than dependency length

Conclusion

- ▶ Annotation **does** influence parsing performance
- ▶ Efficiency – formal properties
- ▶ Accuracy – substantive properties
- ▶ More research needed

- ▶ Giuseppe Attardi. 2006. Experiments with a multilanguage non-projective dependency parser. In *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL)*, pages 166–170.
- ▶ Sabine Buchholz and Erwin Marsi. 2006. CoNLL-X shared task on multilingual dependency parsing. In *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL)*, pages 149–164.
- ▶ Michael Collins. 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Ph.D. thesis, University of Pennsylvania.
- ▶ Sašo Džeroski, Tomaž Erjavec, Nina Ledinek, Petr Pajas, Zdenek Žabokrtsky, and Andreja Žele. 2006. Towards a Slovene dependency treebank. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*.
- ▶ Jan Hajič, Barbora Vidova Hladka, Jarmila Panevová, Eva Hajičová, Petr Sgall, and Petr Pajas. 2001. Prague Dependency Treebank 1.0. LDC, 2001T10.
- ▶ Keith Hall and Vaclav Novák. 2005. Corrective modeling for non-projective dependency parsing. In *Proceedings of the 9th International Workshop on Parsing Technologies (IWPT)*, pages 42–52.
- ▶ Richard A. Hudson. 1990. *English Word Grammar*. Blackwell.
- ▶ Richard Johansson and Pierre Nugues. 2007. Incremental dependency parsing using online learning. In *Proceedings of the CoNLL Shared Task of EMNLP-CoNLL 2007*, pages 1134–1138.
- ▶ Mark Johnson. 1998. PCFG models of linguistic tree representations. *Computational Linguistics*, 24:613–632.
- ▶ Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 423–430.
- ▶ 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, Fernando Pereira, Kiril Ribarov, and Jan Hajič. 2005. Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pages 523–530.
- ▶ Jens Nilsson, Joakim Nivre, and Johan Hall. 2006. Graph transformations in data-driven dependency parsing. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, pages 257–264.
- ▶ Jens Nilsson, Joakim Nivre, and Johan Hall. 2007. Generalizing tree transformations for inductive dependency parsing. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 968–975.
- ▶ Joakim Nivre and Jens Nilsson. 2005. Pseudo-projective dependency parsing. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 99–106.
- ▶ Joakim Nivre, Johan Hall, and Jens Nilsson. 2006a. Maltparser: A data-driven parser-generator for dependency parsing. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, pages 2216–2219.
- ▶ Joakim Nivre, Jens Nilsson, and Johan Hall. 2006b. Talbanken05: A swedish treebank with phrase structure and dependency annotation. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, pages 1392–1395.
- ▶ Joakim Nivre. 2003. An efficient algorithm for projective dependency parsing. In *Proceedings of the 8th International Workshop on Parsing Technologies (IWPT)*, pages 149–160.
- ▶ Joakim Nivre. 2007. Incremental non-projective dependency parsing. In *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL HLT)*, pages 396–403.
- ▶ Joakim Nivre. 2008. Algorithms for deterministic incremental dependency parsing. *Computational Linguistics*, 34:513–553.
- ▶ Joakim Nivre. 2009. Non-projective dependency parsing in expected linear time. In *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics (ACL)*.

- ▶ Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. 2006. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, pages 433–440.
- ▶ Kenji Sagae and Jun'ichi Tsujii. 2008. Shift-reduce dependency DAG parsing. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING)*, pages 753–760.
- ▶ Arnold M. Zwicky. 1985. Heads. *Journal of Linguistics*, 21:1–29.