

PARSEME WG 3

Improving PP attachment in a hybrid dependency parser using semantic, distributional, and lexical resources

EU COST Initiative Meeting
Athens, Greece, March 11-12

Dr. Gerold Schneider

Institute of Computational Linguistics, University of Zurich

English Department, University of Zurich

`gschneid@ifi.uzh.ch`

Q: how much do multi-word resources improve parsing?

1. Multi-Word Terminology

Pro3Gres (Schneider, 2008) uses chunker pre-processing, only parses between heads.

- On in-domain text (Penn, GREVAL):
 - with standard NER (LT-TTT2): worse, most multi-word terms are shorter than chunks
- On out-of-domain (Biomedical):
 - with domain NER: Replace term with term head in pre-processing. Better than chunker, as it corrects many tagging errors (Weeds et al., 2007)
 - with domain-trained tagger: similar to slightly lower performance
 - statistical > lexical resources

2. Improving PP-attachment: Details in Schneider (2012) LREC

Caveat: arguments vs. adjuncts (verbal and nominal):

PP-arguments in Pro3Gres: 90% recall ↔ PP-adjuncts in Pro3Gres: 66% recall

→ are multi-word resources the right tools?

PP attachment relations are multi-word constructions for which many resources exist, and they are highly ambiguous. Our parser uses tri-lexical disambiguation (Collins, 1999)

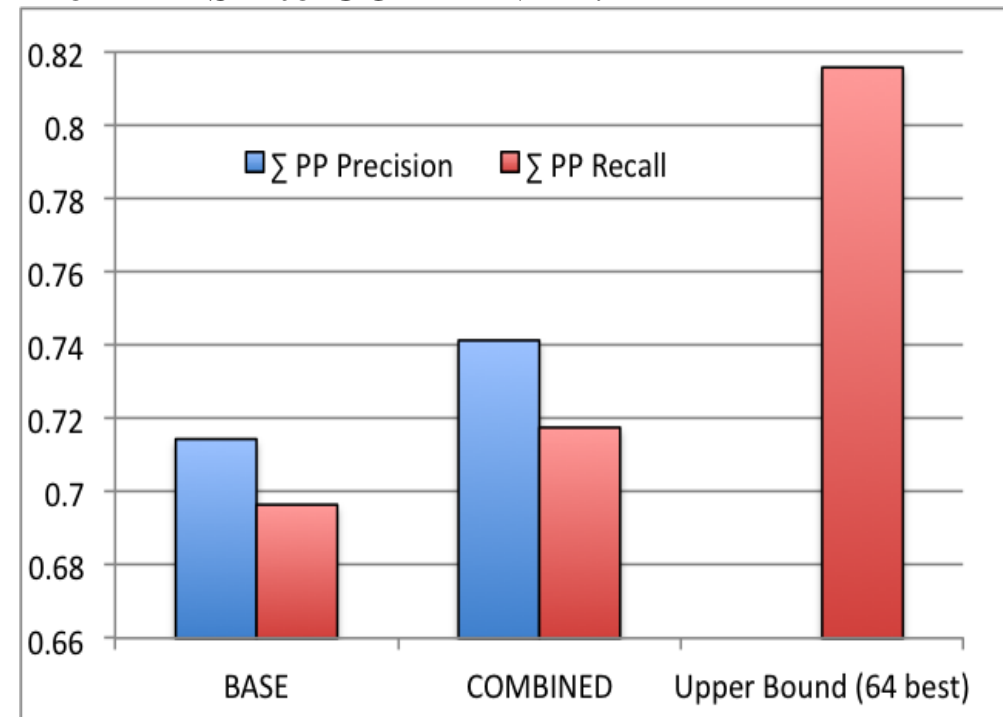
$$p(R, dist|a, b, c) \cong p(R|a, b, c) \cdot p(dist|R) = \frac{f(R, a, b, c)}{f((\sum R), a, b, c)} \cdot \frac{f(R, dist)}{fR}$$

and we added these multi-word resources:

- **semantic expectations** learnt from Penn TB [$p(dog\ hunts) > p(rabbit\ hunts)$]: improves
- **PP interactions** (see DOP: Bod, Scha, and Sima'an (2003)): improves
- **distributional semantics** to alleviate sparse data [$p(preposition_2|v, preposition_1)$], learnt unsupervisedly from the British National corpus (BNC), using non-negative matrix factorisation (Lee and Seung, 2001): marginal improvement
- **self-training**, using the BNC: marginal improvement

- various **lexical resources**, e.g. verb-valency dictionaries: no improvement → implicit in stats

From BASE to COMBINED:



References

- Bod, Rens, Remko Scha, and Khalil Sima'an, editors. 2003. *Data-Oriented Parsing*. Center for the Study of Language and Information, Studies in Computational Linguistics (CSLI-SCL). Chicago University Press.
- Collins, Michael. 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.
- Lee, Daniel D. and H. Sebastian Seung. 2001. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, pages 556–562.
- Schneider, Gerold. 2008. *Hybrid Long-Distance Functional Dependency Parsing*. Doctoral Thesis, Institute of Computational Linguistics, University of Zurich.
- Schneider, Gerold. 2012. Using semantic resources to improve a syntactic dependency parser. In Viktor Pekar Verginica Barbu Mititelu, Octavian Popescu, editor, *SEM-II workshop at LREC 2012*.
- Weeds, Julie, James Dowdall, Gerold Schneider, Bill Keller, and David Weir. 2007. Using distributional similarity to organise BioMedical terminology. In Fidelia Ibekwe-SanJuan, Anne Condamines, and M. Teresa Cabré Castellví, editors, *Application-Driven Terminology Engineering*. Benjamins, Amsterdam/Philadelphia.
-