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

Gerold Schneider, University of Zurich

PP-attachment is a major source of ambiguity for parsers, and discontinuities (words interfering between the governor and the dependent PP) as well as long-distance dependencies (e.g. fronted PP) are frequent. At the same time, noun-PP and particularly verb-PP relations are multi-word constructions for which considerable amounts of resources exist. The dependency parser which we use (Schneider 2008) is hybrid: it combines a hand-written grammar with statistical disambiguation, using a bi-lexical maximum-likelihood model (e.g. Collins 1999) trained on the Penn Treebank. It estimates the probability of the dependency relation *R* at distance (in chunks) *dist*, given the lexical head of the governor (a) and the lexical head of the dependent (b):

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

We improve its performance with a number of statistical multi-word resources:

- WordNet for the bilexical back-off to reduce sparse data: as most words are rare (Zipf 1965), words in combination are particularly rare. We use the WordNet lexicographer file class in the back-off, extending Collins (1995).
- selectional restrictions learnt from the Penn Treebank: we model semantic expectations in addition to syntactic expectations.
- PP-interactions learnt from the Penn Treebank
- distributional semantics to alleviate sparse data, learnt unsupervisedly from the British National corpus (BNC), using non-negative matrix factorisation (Lee and Seung 2001).
- self-training, using the BNC: in order to alleviate the sparse data problem we learn from automatically annotated data. We exploit the fact that back-off level and parser error rate are related, so that attachment decisions which are expected to be unreliable can be overridden by self-trained data, slightly improving performance even without needing a re-ranker (Charniak and Johnson, 2005).

We have also tested various lexical resources, such as verb valency dictionaries, but they did not improve the parser's performance. We hypothesize that argumenthood is already implicitly contained in the statistical frequencies learnt from the Penn Treebank. An analysis of errors also reveals that attachment errors with argument PPs are far fewer than with adjunct PPs.

References

Eugene Charniak and Mark Johnson. 2005. Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 173–180, Ann Arbor, Michigan, June. Association for Computational Linguistics.

Lee, Daniel D. and H. Sebastian Seung. 2001. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, pages 556–562.

Michael Collins and James Brooks. 1995. Prepositional attachment through a backed-off model. In *Proceedings of the Third Workshop on Very Large Corpora*, Cambridge, MA.

Michael Collins. 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.

Gerold Schneider. 2008. *Hybrid Long-Distance Functional Dependency Parsing*. Doctoral Thesis, Institute of Computational Linguistics, University of Zurich.

Zipf, George Kingsley. 1965. *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. The MIT Press, Cambridge, Massachusetts.