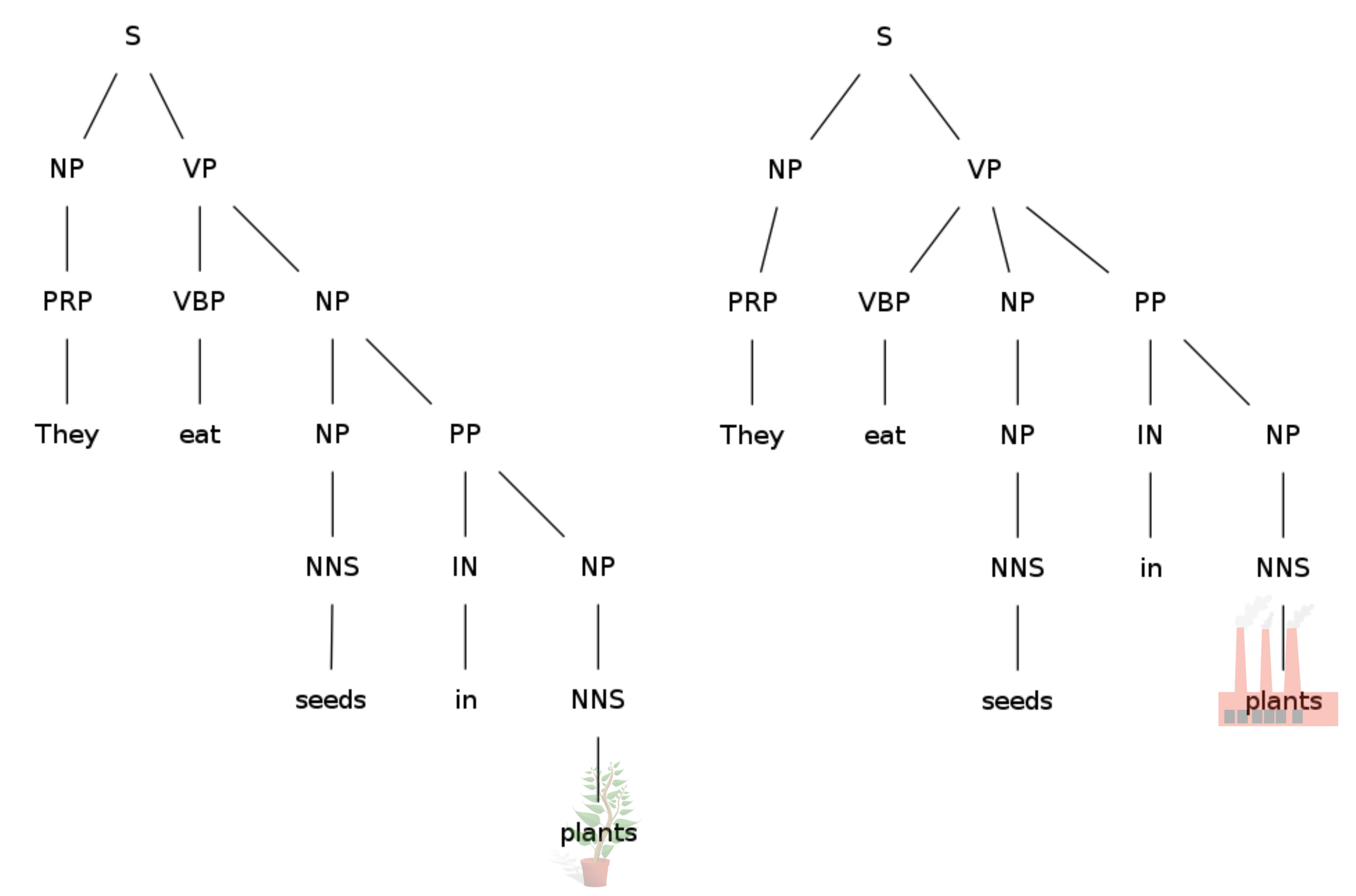
Disambiguating prepositional phrase attachment sites with sense information captured in contextualized distributional data

Clayton Greenberg
Department of Computational Linguistics and Phonetics
Saarland University

Introduction

Determining the governor of a prepositional phrase (PP) is a well-studied parsing problem:



We simplify the problem to a binary classification on quadruples of the form (V, N_1, P, N_2).
In cases such as the above example, the senses of the words in the quadruple interact with the proper attachment classification.
Previous systems have implicitly weighted senses by frequency (Zhao and Lin 2004) or performed explicit word sense disambiguation using WordNet (Stetina and Nagao 1997).
Both of these approaches are problematic because while sense information is crucial, the task of explicit word sense disambiguation (WSD) has lower accuracy than the original task.

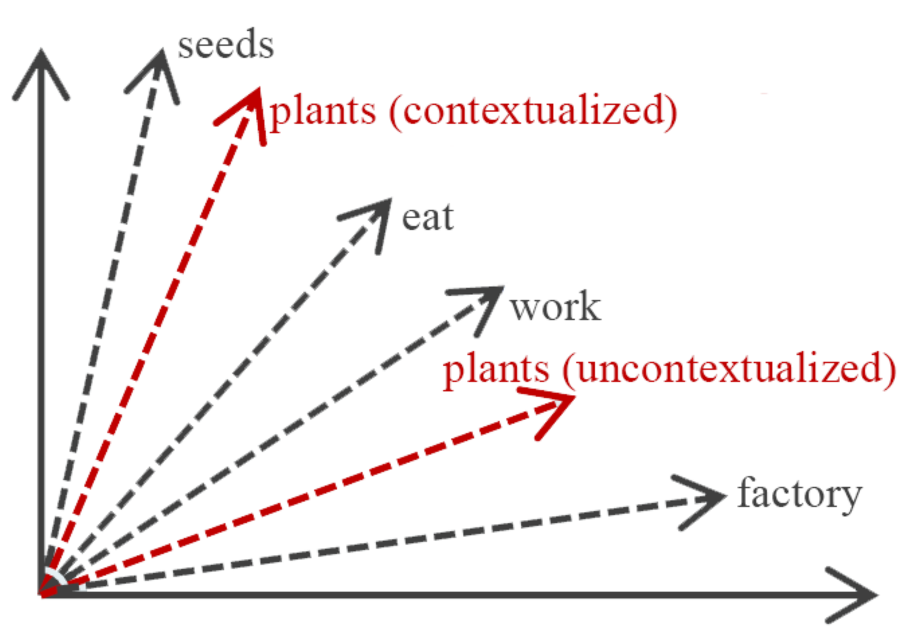
Corpus and model

Ratnaparkhi Reynar and Roukos (1994) extracted 27,937 quadruples from the Penn Treebank.
This corpus has become a standard in the PP-attachment disambiguation literature.
However, many systems preprocessed or edited the corpus, clouding comparability.
For direct comparison with Zhao and Lin (2004), we lemmatize and replace numbers with @.

Dinu and Thater (2012) implemented a vector space model based on the GigaWord corpus.
We use the "filtered" version, so co-occurrences are linked to target words in dependency parses.

The contextualization function is $v(w, c) = \sum_{w' \in W} \alpha(c, w') f(w, w') \vec{e}_{w'}$

where w is the target word, c is the context, W is the set of words, α is the cosine similarity of c and w', f is a co-occurrence function, and $e_{w'}$ is a basis vector.



Classification

We adopt the four step procedure from Zhao and Lin (2004):

1. Consider only the training examples for which all four words are equal to those in the test quadruple.
2. Consider the *k* highest training examples (same preposition) with highest score: $sim(q_1, q_2) = vn_1 + vn_2 + n_1n_2$
3. Same as (2), except using: $sim(q_1, q_2) = v + n_1 + n_2$
4. Assign the default class for the preposition, or noun-attach if there is no default class.

For the word-word similarities, we test four metrics:

abs:	token similarity
noctxt:	uncontextualized vectors
$ctxt_{quad}$:	contextualized vectors using quadruple words only
$ctxt_{sent}$:	contextualized vectors using all words from sentence

Results

Similarity measure	k value	Accuracy
abs	3	80.2%
$noctxt$	11	86.6%
$ctxt_{quad}$	10	**88.4%**
$ctxt_{sent}$	8	81.9%

Step	Coverage	Coverage %	Accuracy
1	244	7.88%	91.8%
2	2849	91.99%	88.1%
3	0	0.00%	N/A
4	4	0.13%	100.0%

Method	Sense handling	Accuracy
BR1994	All senses equal	81.8%
PL2000	Global frequency	84.3%
ZL2004	Global frequency	86.5%
SN1997	Full WSD	88.1%
Our system	Context weighting	**88.4%**
G2013	Full WSD	89.0%

Discussion and conclusion

Improvement over uncontextualized system is statistically significant at $p < 0.04$ level.
Using the full sentence as context is harmful, most likely due to data sparsity.
This system uses only a task-general knowledge base, whereas the highest performing system (Greenberg 2013) requires task-specific resources and labor intensive modifications to training data.

Implicit handling of sense information seems preferable to using frequencies or full WSD.
In future work, we would like to investigate ways to extend the context beyond the quadruple, ideally to whole documents, without losing information as observed here.

References

Georgiana Dinu and Stefan Thater. 2012. Saarland: vector-based models of semantic textual similarity. In *First Joint Conference on Lexical and Computational Semantics (*SEM)*, Montréal. pp. 603–607.

Clayton Greenberg. 2013. *Disambiguating prepositional phrase attachment sites with graded semantic data or, how to rule out elephants in pajamas*. Unpublished undergraduate thesis, Princeton University.

Adwait Ratnaparkhi, Jeff Reynar and Salim Roukos. 1994. A Maximum Entropy Model for Prepositional Phrase Attachment. In *Proceedings of the ARPA Human Language Technology Workshop*, Plainsboro, NJ. pp. 250–255.

Jiri Stetina and Makoto Nagao. 1997. Corpus Based PP Attachment Ambiguity Resolution with a Semantic Dictionary. In *Proceedings of the Fifth Workshop on Very Large Corpora*, Beijing and Hong Kong. pp. 66–80.

Zhao, Shaojun and Dekang Lin. 2004. A Nearest Neighbor Method for Resolving PP-Attachment Ambiguity. In *Proceedings of the First International Joint Conference on Natural Language Processing*, Sanya, China.

Acknowledgments

We would like to acknowledge our anonymous reviewers and Mark Steedman for their suggestions for improving this work. Also, Stefan Thater and Manfred Pinkal for initial advising, and Google for providing travel and conference support.