

Topics on Robust Deep Linguistic Processing

Lexical Acquisition for MWEs, Partial Parse Selection, ...

Yi Zhang

LT-Lab, DFKI GmbH and
Dept of Computational Linguistics
Saarland University
Germany

DELPH-IN Summit 2007
Berlin, Germany

Outline

- 1 Lexical Acquisition for MWEs**
 - Evaluation of the Identification of MWEs
 - Compositional approach vs. “words-with-spaces”
- 2 Partial Parse Selection**
 - Partial parsing with bottom-up chart parser
 - Partial parsing in other processing strategies
- 3 Implementation of Selective Unpacking in PET**

Outline

- 1 Lexical Acquisition for MWEs**
 - Evaluation of the Identification of MWEs
 - Compositional approach vs. “words-with-spaces”
- 2 Partial Parse Selection**
 - Partial parsing with bottom-up chart parser
 - Partial parsing in other processing strategies
- 3 Implementation of Selective Unpacking in PET**

Motivation

Multiword Expressions

Syntactic or semantic properties cannot be derived from their parts [Sag et al., 2002, Villavicencio, 2005]: phrasal verbs, institutionalized phrases, nominal compounds, . . .

Challenge for NLP

It is difficult to provide a unified account for the detection of these distinct but related phenomena.

Grammar Engineering

- Lexical coverage is the major barrier to broad-coverage linguistically deep processing
- MWEs comprise a significant part of the missing lexicon

Previous Work [Zhang et al., 2006]

- Error-mining to detect problematic n-grams
 - Statistical measures to detect genuine MWEs
 - New MWE entries are created with automated lexical acquisition
 - *Words-with-spaces* approach
- + Parser coverage is improved significantly
- Validation steps are not thoroughly evaluated
 - Parser accuracy is not investigated

Parser Coverage

	Entries Added	Item #	Covered #	Coverage
ERG	–	6246	268	4.3%
ERG+MWE	373	6246	1168	18.7%

Identification of MWEs

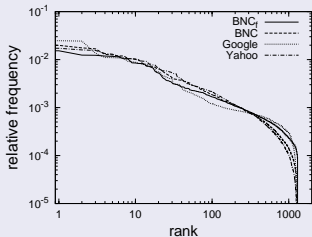
Questions

- How reliable is the corpus used?
- How precise is a statistical measure to distinguish the phenomena studied?

Comparing Corpora

Hypothesis

The relative ordering in frequency for different n-grams is preserved across corpora, in the same domain. If not, different conclusions may be drawn from different corpora



	BNC	Google	Yahoo
BNC_f	0.81	0.73	0.78
BNC	–	0.73	0.77
Google	–	–	0.86

- The overall ranking distribution is very similar for these corpora, showing the expected Zipf like behavior

Comparing Statistical Measures

Are They Equivalent?

	$MI \times \chi^2$	$MI \times PE$	$\chi^2 \times PE$
Q	0.71	0.55	0.45

- Q is the probability of finding the same ordering for two trigrams
- 70% chance of getting the same order from MI and χ^2 , they are very different from the PE
- MI or PE seem to differentiate between MWEs and non-MWEs

Compositional approach vs. "words-with-spaces"

Lexical Acquisition for MWE Heads

- Rank all the MWE candidates according to the three statistical measures: MI, χ^2 , PE, and select the top 30 MWE with highest average ranking
- Extract sub-corpus from BNC_f which contains at least one of the MWE for evaluation (674 sentences)
- Use heuristics to extract head words (20 head words)
- Run lexical acquisition for head words on the sub-corpus (21 new entries)

Compositional approach vs. "words-with-spaces"

Grammar Coverage

	item #	parsed #	avg. analysis #	coverage %
ERG	674	48	335.08	7.1%
ERG + MWE	674	153	285.01	22.7%

- The coverage improvement is largely compatible with the results of "words-with-spaces" approach reported in [Zhang et al., 2006] (about 14%)
- Great reduction in lexical entries added

Compositional approach vs. "words-with-spaces"

Grammar Accuracy

- 153 parsed sentences are analyzed by hand
- 124 (81.0%) of them receive at least one correct/acceptable analysis (comparable to the accuracy reported by [Baldwin et al., 2004])
- Parse selection model finds best analysis in top-5 for 66% of the cases, and top-10 for 75%

Outline

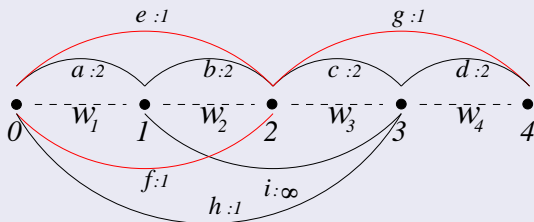
- 1 **Lexical Acquisition for MWEs**
 - Evaluation of the Identification of MWEs
 - Compositional approach vs. “words-with-spaces”
- 2 **Partial Parse Selection**
 - Partial parsing with bottom-up chart parser
 - Partial parsing in other processing strategies
- 3 **Implementation of Selective Unpacking in PET**

Partial Parsing

- Full analysis are not always available
 - limited grammar coverage
 - ungrammatical/extra-grammatical inputs
- All-or-Nothing
- Partial analyses are available as intermediate parsing results
 - e.g. passive edges in bottom-up chart parsing
- Not all intermediate parsing results are good partial analysis candidates
- Proper selection model is the key

Partial parsing with bottom-up chart parser

Shortest Path [Kasper et al., 1999]



- A partial analysis is defined as a set of passive edges that connects a path that covers the entire input sequence
- Each edge is heuristically weighted
- The shortest path is the best partial analyses
 - Heuristic weights are less accurate and not trainable
 - It does not fully disambiguate

Partial parsing with bottom-up chart parser

More Sophisticated Partial Parse Ranking

$$P(\Phi|w) \approx P(\Omega|w) \cdot \prod_{i=1}^k P(t_i|w_i)$$

- Segmentation model: $P(\Omega|w)$
- Parse ranking (ME) model: $P(t_i|w_i) = \frac{\exp \sum_{j=1}^n \lambda_j f_j(t_i, w_i)}{\sum_{t' \in T} \exp \sum_{j=1}^n \lambda_j f_j(t', w_i)}$

Evaluation

- Syntactic structure: 38.5% \rightarrow 51.7%
- Semantics (RMRS, distance): 0.674 \rightarrow 0.296

Partial Parsing in Other Processing Strategies

- Passive edges in bottom-up chart parsing only provide local analyses; Long distance dependencies beyond edge span are usually lost
- To combine with other (top-down?) processing strategies to recover long distance dependencies

Outline

- 1 Lexical Acquisition for MWEs**
 - Evaluation of the Identification of MWEs
 - Compositional approach vs. “words-with-spaces”
- 2 Partial Parse Selection**
 - Partial parsing with bottom-up chart parser
 - Partial parsing in other processing strategies
- 3 Implementation of Selective Unpacking in PET**

What Is It?

Ambiguity Packing [Oepen and Carroll, 2000]

- Implemented in LKB and PET
- Parse forest packed with subsumption relation are not guaranteed to be consistent (among packed daughters)
- The read-out procedure needs to replay “unifications”

Selective Unpacking [Carroll and Oepen, 2005]

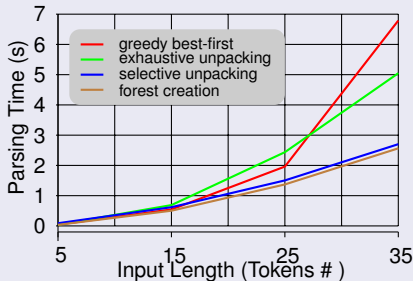
- Minimizes read-out cost by extracting the best readings first
- Unification operations are delayed and applied only on best top level hypotheses

Extensions [Zhang et al., 2007]

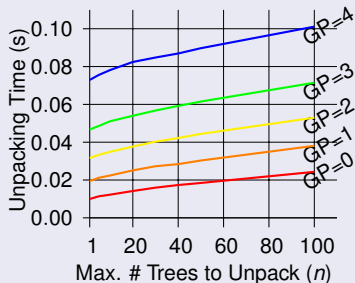
- Allow grandparenting features to improve disambiguation accuracy
- Instantiation failure caching and propagation to optimize performance on highly inconsistent parse forests

Empirical Results

Parsing Time to i -length



Unpacking Time to n



Instantiation failure caching and propagation with GG

Configuration	Unifications (#)	Hypotheses (#)	Unpack (ms)	Total (ms)
no caching	5535	1245	70	410
with caching	4915	382	10	350

How To Use It?

- Without ambiguity packing

```
$ cheap [-packing=0] ...
```

- Retroactive ambiguity packing with exhaustive unpacking

```
$ cheap -packing [-nsolutions=0] ...
```

- Ambiguity packing with selective unpacking

```
$ cheap -packing -nsolutions=10 ...
```

<http://wiki.delph-in.net/moin/PetSelectiveUnpacking>

Summary

- MWE acquisition experiment shows that a compositional approach helps to improve parser coverage without compromising accuracy
- A better partial parse selection model is necessary for robust processing
- Selective unpacking brings significant efficiency (and accuracy) gain

Thanks!

For Further Reading I



Baldwin, T., Bender, E. M., Flickinger, D., Kim, A., and Oepen, S. (2004).

Road-testing the English Resource Grammar over the British National Corpus.

In Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal.



Carroll, J. and Oepen, S. (2005).

High efficiency realization for a wide-coverage unification grammar.

In Proceedings of the 2nd International Joint Conference on Natural Language Processing (IJCNLP 2005), pages 165–176, Jeju Island, Korea.



Kasper, W., Kiefer, B., Krieger, H.-U., Rupp, C., and Worm, K. (1999).

Charting the depths of robust speech processing.

In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL 1999), pages 405–412, Maryland, USA.



Oepen, S. and Carroll, J. (2000).

Ambiguity packing in constraint-based parsing — practical results.

In Proceedings of the 1st Annual Meeting of the North American Chapter of Association for Computational Linguistics (NAACL 2000), pages 162–169, Seattle, USA.



Sag, I., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. (2002).

Multiword expressions: a pain in the neck for NLP.

In Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002), pages 1–15, Mexico City, Mexico.

For Further Reading II



Villavicencio, A. (2005).

The availability of verb-particle constructions in lexical resources: how much is enough?
Journal of Computer Speech and Language Processing, 19.



Zhang, Y., Oepen, S., and Carroll, J. (2007).

Efficiency in unification-based N-best parsing.
In Proceedings of the 10th International Conference on Parsing Technologies (IWPT 2007), pages 48–59,
Prague, Czech.



Zhang, Y., Villavicencio, A., Kordoni, V., and Idiart, M. (2006).

Automated multiword expression prediction for grammar engineering.
In Proceedings of the ACL 2006 Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties, pages 36–44, Sydney, Australia.