Semantic Feature Engineering for Enhancing Disambiguation Performance in Deep Linguistic Processing

Danielle Ben-Gera♣, Yi Zhang♢, Valia Kordoni♢
{danielle,yzhang,kordoni}@coli.uni-sb.de

♣ Dept. of Computational Linguistics (COLI), Saarland University
♢ German Research Centre for Artificial Intelligence (DFKI GmbH)

LREC 2010

Outline

1. **Motivation**
2. **Experiments**
   - Baseline
   - Deep semantic features
3. **Results**
4. **Conclusion**

Outline

1. **Motivation**
2. Experiments
   - Baseline
   - Deep semantic features
3. Results
4. Conclusion

Motivation

Fine-grained deep grammars

- Wide and meaningful coverage.
- Many uses in NLP:
  - Machine Translation
  - Question Answering
  - ...

→ But: often license a vast number of structures that make the usage of those grammars difficult.

Solution: Parse disambiguation

Using statistical approaches to train a model in order to rank the different parses.

Motivation

Fine-grained deep grammars

- Wide and meaningful coverage.
- Many uses in NLP:
  - Machine Translation
  - Question Answering
  - ...

→ But: often license a vast number of structures that make the usage of those grammars difficult.

Solution: Parse disambiguation

Using statistical approaches to train a model in order to rank the different parses.

Motivation

Fine-grained deep grammars

- Wide and meaningful coverage.
- Many uses in NLP:
  - Machine Translation
  - Question Answering
  - ...

$\rightarrow$ But: often license a vast number of structures that make the usage of those grammars difficult.

Solution: Parse disambiguation

Using statistical approaches to train a model in order to rank the different parses.

Approaches

Generative methods

Probabilistic parsing; PCFG like derivations

- Early pruning.
- Difficult to integrate non-local features.
- Independence assumption between features.
- Inflexible: hard to integrate new features.

[Magerman, 1995], [Collins, 1997], [Charniak, 1997], [Roark, 2001], ..

Approaches (2)

Discriminative methods

Ranking parses; Log Linear models:

- Re-ranking of the parser's output.
- Easy integration of new features.
- No independence assumption.

[Charniak, 2000], [Riezler et al., 2002], [Toutanova et al., 2005], [Collins and Koo, 2005], [Fujita et al., 2007]

Outline

1 Motivation

2 **Experiments**
- Baseline
- Deep semantic features

3 Results

4 Conclusion

The Setup

Delph-in Colloboration

Set of tools and Grammars for NLP.

- Available at http://www.delph-in.net.

Our Framework

- The Datasets: *LOGON* and *WeScience* Treebanks.
  - The Grammar: HPSG English Resource Grammar (Lingo ERG).
  - The Parser: PET parser for unification-based grammars.
  - Contains deep syntactic and semantic information.
- The Classifier: Maximum Entropy classifier
  - *TADM* - Toolkit for Advanced Discriminative Models.

Choosing the Baseline

Informative Baseline

- Should allow comparison with other approaches:
  - Common practice to choose syntactic features ([Toutanova et al., 2005], [Fujita et al., 2007], [Zhang et al., 2007]).
- Should provide a good testing measure for our approach:
  - We are testing the effects of adding semantic information.

→Using syntactic elements only, incorporating non-local features.

Choosing the Baseline

Informative Baseline

- Should allow comparison with other approaches:
  - Common practice to choose syntactic features ([Toutanova et al., 2005], [Fujita et al., 2007], [Zhang et al., 2007]).
- Should provide a good testing measure for our approach:
  - We are testing the effects of adding semantic information.

→Using syntactic elements only, incorporating non-local features.

Baseline: Results

In Domain Results:

	features	LOGON		WeScience	
		1-best	10-best	1-best	10-best
p0	233,982	49.2154	75.1783	40.2282	69.0442
p1	349,564	54.0656	78.8873	43.2239	71.6119
p2	1,008,198	54.3509	77.4607	46.7902	74.7503
p3	2,493,884	55.7774	79.7432	49.2154	75.1783

Domain Adaptation Results:

	features	WS-LO		LO-WS	
		1-best	10-best	1-best	10-best
p0	233,982	31,6690	62.9101	27.1041	56.4907
p1	349,564	32.6676	63.7660	29.5292	62.9201
p2	1,008,198	35.0927	67.0470	30.2442	59.7717
p3	2,493,884	34.0941	66.9044	31.5263	63.4807

Semantic Modules

Minimal Recursion Semantics

- Fully underspecified flat semantics.
- Captures ambiguities.
- Highly Expressive.
- Can be easily incorporated into the constraint based HPSG.

Elementary Dependency Structures

- Shallow Dependency structures
- Captures basic relations between words, particularly predicate-argument relations (similar to an MRS solved form).
- Can be automatically extracted from the MRS.

Semantic Modules

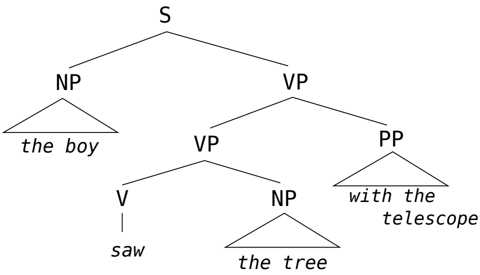
Minimal Recursion Semantics

- Fully underspecified flat semantics.
- Captures ambiguities.
- Highly Expressive.
- Can be easily incorporated into the constraint based HPSG.

Elementary Dependency Structures

- Shallow Dependency structures
- Captures basic relations between words, particularly predicate-argument relations (similar to an MRS solved form).
- Can be automatically extracted from the MRS.

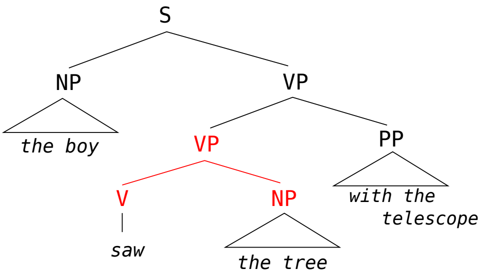
Example



Example

Syntactic Features:

Example

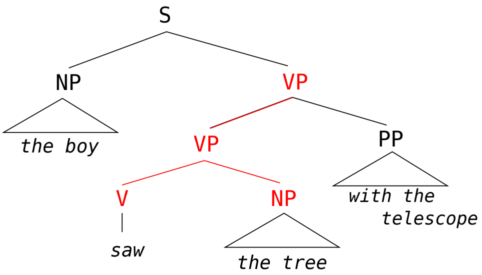


Example

Syntactic Features:

<syn:p0>vp:v,np

Example

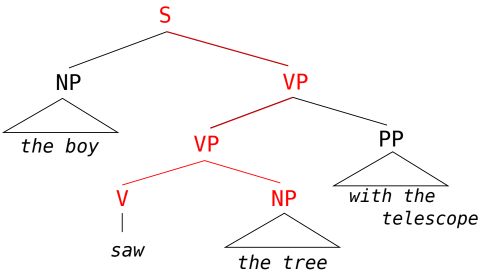


Example

Syntactic Features:

<syn:p0>vp:v,np
<syn:p1>vp,vp:v,np

Example

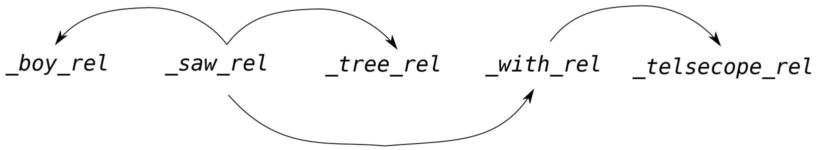


Example

Syntactic Features:

<syn:p0>vp:v,np
<syn:p1>vp,vp:v,np
<syn:p2>s,vp,vp:v,np
. . .

Example



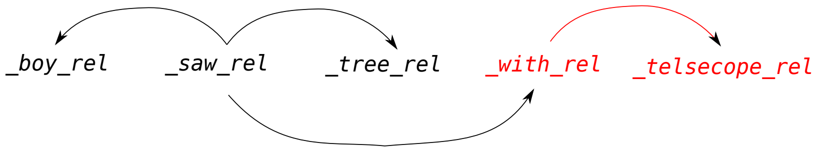
Example

Syntactic Features:

<syn:p0>vp:v,np
<syn:p1>vp,vp:v,np
<syn:p2>s,vp,vp:v,np
. . .

Semantic Features:

Example



Example

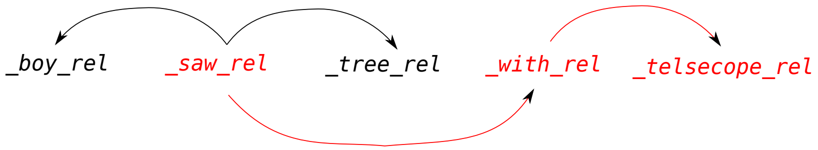
Syntactic Features:

<syn:p0>vp:v,np
<syn:p1>vp,vp:v,np
<syn:p2>s,vp,vp:v,np
. . .

Semantic Features:

<sem:d1>with:telescope

Example



Example

Syntactic Features:	Semantic Features:
<syn:p0>vp:v,np	<sem:d1>with:telescope
<syn:p1>vp,vp:v,np	<sem:d2>saw,with:telescope
<syn:p2>s,vp,vp:v,np	. . .
. . .	. . .

Outline

1. Motivation
2. Experiments
   - Baseline
   - Deep semantic features
3. Results
4. Conclusion

Semantics Models

In Domain Results:

Model	# features	LOGON		WeScience	
		1-best	10-best	1-best	10-best
Random Pick		14.7113	37.0314	13.5550	35.0914
sem-eds	1,265,442	31.8116	67.4750	18.9728	50.9272
sem-mrs	159,420	37.8031	72.1825	25.3922	58.2025
sem-combined	1,424,862	42.5106	76.4621	28.1027	64.6219

Domain Adaptation Results:

Model	# features	WS-LO		LO-WS	
		1-best	10-best	1-best	10-best
Random Pick		14.7113	37.0314	13.5550	35.0914
sem-eds	1,265,442	14.4079	42.7960	12.6961	40.9415
sem-mrs	159,420	21.2553	53.9229	17.4037	49.3580
sem-combined	1,424,862	25.2496	56.2054	18.5449	52.6390

Combined Models

In Domain Results:

	features	LOGON 1-best	LOGON 10-best	WeScience 1-best	WeScience 10-best
syn:p3	2,493,884	55.7774	79.7432	49.2154	75.1783
sem:mrs+eds	2,736,573	42.5106	76.4621	28.1027	64.6219
syn+sem	5,230,457	59.6291	82.0256	47.3609	75.3209

Domain Adaptation Results:

	features	WS-LO 1-best	WS-LO 10-best	LO-WS 1-best	LO-WS 10-best
syn:p3	2,493,884	33.3808	64.1949	31.5263	63.4807
sem:mrs+eds	2,736,573	25.2496	56.2054	18.2596	52.4964
syn+sem	5,230,457	36.9472	68.1883	29.6718	62.3395

Outline

1 Motivation

2 Experiments
- Baseline
- Deep semantic features

3 Results

4 Conclusion

Conclusion

- Syntactic features have reached their limit:
  $\Rightarrow$ adding semantic information.
- MRS information performs very well with a small set of features.
- Using different data-sets might influence the results.

Charniak, E. (1997).
Statistical parsing with a context-free grammar and word statistics.
In *Proceedings of the National Conference on Artificial Intelligence*, pages 598–603.

Charniak, E. (2000).
A maximum-entropy-inspired parser.
In *Proceedings of the 6th Applied Natural Language Processing Conference (ANLP 2000)*.

Collins, M. (1997).
Three generative, lexicalised models for statistical parsing.
In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 16–23.

Collins, M. and Koo, T. (2005).
Discriminative reranking for natural language parsing.

Computational Linguistics, 31(1):25–70.

Fujita, S., Bond, F., Oepen, S., and Tanaka, T. (2007).
Exploiting semantic information for HPSG parse selection.
ACL 2007, 100.

Huang, L. (2008).
Forest reranking: Discriminative parsing with non-local features.
In *Proceeding of Association for Computational Linguistics (ACL 2008)*.

Magerman, D. M. (1995).
Statistical decision-tree models for parsing.
In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, pages 276–283.

Malouf, R. and van Noord, G. (2004).

Wide coverage parsing with stochastic attribute value grammars.
In *In Proceedings of the IJCNLP-04 workshop: beyond shallow analyses - formalisms and statistical modeling for deep analyses*.

Riezler, S., King, T., Kaplan, R., Crouch, R., Maxwell III, J., and Johnson, M. (2002).
Parsing the Wall Street Journal using a Lexical-Functional Grammar and discriminative estimation techniques.
In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 271–278.

Roark, B. (2001).
Probabilistic top-down parsing and language modeling.
Computational Linguistics, 27(2):249–276.

Toutanova, K., Manning, C., Flickinger, D., and Oepen, S. (2005).

Stochastic HPSG parse disambiguation using the Redwoods corpus.
Research on Language & Computation, 3(1):83–105.

Zhang, Y., Oepen, S., and Carroll, J. (2007).
Efficiency in unification-based n-best parsing.
IWPT, pages 48–57.