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
  - ...

→ 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 Collaboration**
Set of tools and Grammars for NLP.
- Available at http://www.delph-in.net.

**Our Framework**
- The Datasets: LOGON and WeScience Treebanks.
  - 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.
# Motivation

# Experiments

# Results

## Baseline

### Baseline: Results

#### In Domain Results:

<table>
<thead>
<tr>
<th></th>
<th>features</th>
<th>LOGON</th>
<th></th>
<th>WeScience</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1-best</td>
<td>10-best</td>
<td>1-best</td>
<td>10-best</td>
</tr>
<tr>
<td>p0</td>
<td>233,982</td>
<td>49.2154</td>
<td>75.1783</td>
<td>40.2282</td>
<td>69.0442</td>
</tr>
<tr>
<td>p1</td>
<td>349,564</td>
<td>54.0656</td>
<td>78.8873</td>
<td>43.2239</td>
<td>71.6119</td>
</tr>
<tr>
<td>p2</td>
<td>1,008,198</td>
<td>54.3509</td>
<td>77.4607</td>
<td>46.7902</td>
<td>74.7503</td>
</tr>
<tr>
<td>p3</td>
<td>2,493,884</td>
<td>55.7774</td>
<td>79.7432</td>
<td>49.2154</td>
<td>75.1783</td>
</tr>
</tbody>
</table>

#### Domain Adaptation Results:

<table>
<thead>
<tr>
<th></th>
<th>features</th>
<th>WS-LO</th>
<th></th>
<th>LO-WS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1-best</td>
<td>10-best</td>
<td>1-best</td>
<td>10-best</td>
</tr>
<tr>
<td>p0</td>
<td>233,982</td>
<td>31.6690</td>
<td>62.9101</td>
<td>27.1041</td>
<td>56.4907</td>
</tr>
<tr>
<td>p1</td>
<td>349,564</td>
<td>32.6676</td>
<td>63.7660</td>
<td>29.5292</td>
<td>62.9201</td>
</tr>
<tr>
<td>p2</td>
<td>1,008,198</td>
<td>35.0927</td>
<td>67.0470</td>
<td>30.2442</td>
<td>59.7717</td>
</tr>
<tr>
<td>p3</td>
<td>2,493,884</td>
<td>34.0941</td>
<td>66.9044</td>
<td>31.5263</td>
<td>63.4807</td>
</tr>
</tbody>
</table>
Deep semantic features

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.
Deep semantic features

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.
Deep semantic features

Example

```
the boy
  V
  | saw
 VP
 NP
  with the
telescope

S
 NP
 VP
 PP
```

Syntactic Features:
Deep semantic features

Example

```
the boy

VP
  V
  saw

NP
  the tree

PP
  with the telescope
```

Syntactic Features:

\(<\text{syn:p0}>\text{vp:v},\text{np}\)
Deep semantic features

Example

Syntactic Features:

<syn:p0>vp:v,np
<syn:p1>vp, vp:v, np
Deep semantic features

Example

S
  NP  VP
  the boy  V  NP
  VP  PP
  saw  with the
  NP  the telescope

Syntactic Features:

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

Example

Syntactic Features:

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

...
Deep semantic features

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`
Deep semantic features

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 . . .
. . .
# Semantics Models

## In Domain Results:

<table>
<thead>
<tr>
<th>Model</th>
<th># features</th>
<th>LOGON</th>
<th>WeScience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Pick</td>
<td></td>
<td>14.7113</td>
<td>13.5550</td>
</tr>
<tr>
<td></td>
<td></td>
<td>37.0314</td>
<td>35.0914</td>
</tr>
<tr>
<td>sem-eds</td>
<td>1,265,442</td>
<td>31.8116</td>
<td>18.9728</td>
</tr>
<tr>
<td></td>
<td></td>
<td>67.4750</td>
<td>50.9272</td>
</tr>
<tr>
<td>sem-mrs</td>
<td>159,420</td>
<td>37.8031</td>
<td>25.3922</td>
</tr>
<tr>
<td></td>
<td></td>
<td>72.1825</td>
<td>58.2025</td>
</tr>
<tr>
<td>sem-combined</td>
<td>1,424,862</td>
<td>42.5106</td>
<td>28.1027</td>
</tr>
<tr>
<td></td>
<td></td>
<td>76.4621</td>
<td>64.6219</td>
</tr>
</tbody>
</table>

## Domain Adaptation Results:

<table>
<thead>
<tr>
<th>Model</th>
<th># features</th>
<th>WS-LO</th>
<th>LO-WS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Pick</td>
<td></td>
<td>14.7113</td>
<td>13.5550</td>
</tr>
<tr>
<td></td>
<td></td>
<td>37.0314</td>
<td>35.0914</td>
</tr>
<tr>
<td>sem-eds</td>
<td>1,265,442</td>
<td>14.4079</td>
<td>12.6961</td>
</tr>
<tr>
<td></td>
<td></td>
<td>42.7960</td>
<td>40.9415</td>
</tr>
<tr>
<td>sem-mrs</td>
<td>159,420</td>
<td>21.2553</td>
<td>17.4037</td>
</tr>
<tr>
<td></td>
<td></td>
<td>53.9229</td>
<td>49.3580</td>
</tr>
<tr>
<td>sem-combined</td>
<td>1,424,862</td>
<td>25.2496</td>
<td>18.5449</td>
</tr>
<tr>
<td></td>
<td></td>
<td>56.2054</td>
<td>52.6390</td>
</tr>
</tbody>
</table>
## Combined Models

### In Domain Results:

<table>
<thead>
<tr>
<th>features</th>
<th>LOGON 1-best</th>
<th>LOGON 10-best</th>
<th>WeScience 1-best</th>
<th>WeScience 10-best</th>
</tr>
</thead>
<tbody>
<tr>
<td>syn:p3</td>
<td>55.7774</td>
<td>79.7432</td>
<td>49.2154</td>
<td>75.1783</td>
</tr>
<tr>
<td>sem:mrs+eds</td>
<td>42.5106</td>
<td>76.4621</td>
<td>28.1027</td>
<td>64.6219</td>
</tr>
<tr>
<td>syn+sem</td>
<td>59.6291</td>
<td>82.0256</td>
<td>47.3609</td>
<td>75.3209</td>
</tr>
</tbody>
</table>

### Domain Adaptation Results:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>syn:p3</td>
<td>33.3808</td>
<td>64.1949</td>
<td>31.5263</td>
<td>63.4807</td>
</tr>
<tr>
<td>sem:mrs+eds</td>
<td>25.2496</td>
<td>56.2054</td>
<td>18.2596</td>
<td>52.4964</td>
</tr>
<tr>
<td>syn+sem</td>
<td>36.9472</td>
<td>68.1883</td>
<td>29.6718</td>
<td>62.3395</td>
</tr>
</tbody>
</table>
Outline

1. Motivation

2. Experiments
   - Baseline
   - Deep semantic features

3. Results

4. Conclusion
Syntactic features have reached their limit:  
⇒ adding semantic information.

MRS information performs very well with a small set of features.

Using different data-sets might influence the results.


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.*

Parsing the Wall Street Journal using a Lexical-Functional Grammar and discriminative estimation techniques.


Probabilistic top-down parsing and language modeling.


Stochastic HPSG parse disambiguation using the Redwoods corpus.  
