Identifying Semantic Roles
Using Combinatory Categorial Grammar

Daniel Gildea and Julia Hockenmaier
University of Pennsylvania
Introduction

Understanding difficult due to variation in syntactic realization of semantic roles:

- John will meet with Mary.
- John will meet Mary.
- John and Mary will meet.
- The door opened.
- Mary opened the door.
Statistical Approaches to Semantic Roles

Gildea & Palmer ACL 2002: Predict PropBank roles using features derived from Treebank parser output (Collins).

Similar approaches:

- FrameNet data: Gildea & Jurafsky 2000

Problem: Long distance dependencies difficult to find/interpret.
Long Distance Dependencies

Standard Treebank parsers do not return dependencies from relative clauses, *wh*-movement, control, raising.

truth: \[ \text{[ARG0 Big investment banks] refused to step up to the plate to support [ARG1 the floor traders]} \].

system: Big investment banks refused to step up to the plate to support [ARG1 the floor traders].

CCG parsers return local and long-distance dependencies in same form.
Overview

• Semantic roles in PropBank

• Combinatory Categorial Grammar

• Features: matching CCG and PropBank

• Results and Discussion
PropBank

• Role labels defined per-predicate:
  – Core: Arg0, Arg1, ...
  – ArgM: Temporal, Locative, etc

• Rolesets correspond to senses

• Tagging all verbs in treebanked Wall Street Journal

• Preliminary corpus: 72,109 verb instances (2462 unique verbs), 190,815 individual arguments (75% are “core”)

Kingsbury et al., HLT 2002
Sample PropBank Roleset Entry

**offer**  
Arg0: entity offering  
Arg1: commodity  
Arg2: benefactive or entity offered to  
Arg3: price

- [Arg0 the company] to **offer** [Arg1 a 15% stake] to [Arg2 the public].
- [Arg0 Sotheby’s] ... **offered** [Arg2 the Dorrance heirs] [Arg1 a money-back guarantee]
PropBank ArgM Roles

Location, Time, Manner, Direction, Cause, Discourse, Extent, Purpose, Negation, Modal, Adverbial

- Location: *in Tokyo*
- Discourse: *However*
- Negation: *not*
Probability Model for Predicting Roles

Based on features extracted from parser output:

- **Phrase type**: NP, PP, S, etc
- **Position**: Before/after predicate word
- **Voice**: Active/passive
- **Head Word**: Uses head rules of parser
- **Parse Tree Path**: syntactic relation to predicate

*Gildea and Palmer ACL 2002*
Parse Tree Path

Ex: \( P(fe|p = “eat”, path = “VB↑VP↑S↓NP”, head = “He”) \)
Backoff Lattice
Sentence-Level Argument Assignment

Choose best assignment of roles $r_{1..n}$ given predicate $p$, and features $F_{1..n}$:

$$P(r_{1..n}|F_{1..n}, p) \approx P(\{r_{1..n}\}|p) \prod_i \frac{P(r_i|F_i, p)}{P(r_i|p)}$$

Argument set probabilities provide (limited) dependence between individual labeling decisions.
Combinatory Categorial Grammar

- **Categories** specify subcat lists of words/constituents
  
  Declarative verb phrase: \( S[\text{dcl}] \backslash NP \)
  
  Transitive declarative verb: \( (S[\text{dcl}] \backslash NP) / NP \)

- **Combinatory rules** specify how constituents can combine.

- **Derivations** spell out process of combining constituents

\[
S[\text{dcl}]
\]
\[
\begin{array}{c}
\text{NP} \\
\text{London} \\
denied \\
\end{array}
\begin{array}{c}
S[\text{dcl}] \backslash NP \\
(S[\text{dcl}] \backslash NP) / NP \\
\text{plans} \\
\end{array}
\]

Gildea & Hockenmaier  EMNLP 2003 13
Predicate-argument structure in CCG

- The **argument slots** of functor categories define dependencies:

$$S[\text{dcl}]$$

- $$NP_1$$
  - London
  - $$(S[\text{dcl}] \backslash NP_1) / NP_2$$
  - denied
- $$NP_2$$
  - plans
Long-range dependencies in CCG

- **Long-range dependencies** are projected from the lexicon:

```
NP
    \ NP / NP_2
    |    |
plans (NP \ NP_i) / (S[dcl] / NP_i)
    |    |
that (S / (S \ NP_1) / (S[dcl] \ NP_1) / NP_2)
    |    |
NP / denied
    |
London
```

- Similar for control, raising, etc.
## CCG Predicate-Argument Relations

London denied plans on Monday

<table>
<thead>
<tr>
<th>$w_h$</th>
<th>$w_a$</th>
<th>$c_h$</th>
<th>$i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>denied</td>
<td>London</td>
<td>(S[dcl]\text{NP}_1)/\text{NP}_2</td>
<td>1</td>
</tr>
<tr>
<td>denied</td>
<td>plans</td>
<td>(S[dcl]\text{NP}_1)/\text{NP}_2</td>
<td>2</td>
</tr>
<tr>
<td>on</td>
<td>denied</td>
<td>(((S\text{NP}_1)\text{NP}_2))/\text{NP}_3</td>
<td>2</td>
</tr>
<tr>
<td>on</td>
<td>Monday</td>
<td>(((S\text{NP}_1)\text{NP}_2))/\text{NP}_3</td>
<td>3</td>
</tr>
</tbody>
</table>

Gildea & Hockenmaier  
EMNLP 2003
CCG and PropBank

- CCG derivation often doesn’t match Penn Treebank constituent structure

- Training: Find maximal projection in CCG of headword of constituent labeled in PropBank

- Evaluation: Score on headwords, rather than constituent boundaries
Mismatches between CCGbank and PropBank

- 23% of PropBank arguments do not correspond to CCG relations:
  - to *offer* ...[PP to [NP_{ARG2} the public]]

We use a path feature instead:

```
S[b] \ NP
```

```
S[b] \ NP
```

```
((S\NP) \ (S\NP)) / NP
```

```
NP
```

```
to
```

```
the public
```

Sparser than Treebank path feature.
Experiment

Train on Sections 02-21, test on 23.

- Compare Treebank- and CCG-based systems
- Compare automatic parser output and gold standard parses
- Compare Treebank parses with and without traces
## Accuracy of Semantic Role Prediction

<table>
<thead>
<tr>
<th>Parsed</th>
<th></th>
<th>Treebank-based</th>
<th></th>
<th>CCG-based</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Used</td>
<td>Arg</td>
<td>Prec</td>
<td>Recall</td>
<td>F-score</td>
</tr>
<tr>
<td></td>
<td>Automatic</td>
<td>core</td>
<td>75.9</td>
<td>69.6</td>
<td>72.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>72.6</td>
<td>61.2</td>
<td>66.4</td>
</tr>
<tr>
<td></td>
<td>Gold-standard</td>
<td>core</td>
<td>85.5</td>
<td>81.7</td>
<td>83.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>78.8</td>
<td>69.9</td>
<td>74.1</td>
</tr>
<tr>
<td></td>
<td>Gold-standard</td>
<td>core</td>
<td>77.6</td>
<td>75.2</td>
<td>76.3</td>
</tr>
<tr>
<td></td>
<td>w/o traces</td>
<td>all</td>
<td>74.4</td>
<td>66.5</td>
<td>70.2</td>
</tr>
</tbody>
</table>
## Comparison of scoring regimes

<table>
<thead>
<tr>
<th>Parses</th>
<th>Scoring</th>
<th>Treebank-based</th>
<th></th>
<th></th>
<th>CCG-based</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Prec</td>
<td>Recall</td>
<td>F-score</td>
<td>Prec</td>
<td>Recall</td>
<td>F-score</td>
</tr>
<tr>
<td>Automatic</td>
<td>Head word</td>
<td>72.6</td>
<td>61.2</td>
<td>66.4</td>
<td>71.0</td>
<td>63.1</td>
<td>66.8</td>
</tr>
<tr>
<td></td>
<td>Boundary</td>
<td>68.6</td>
<td>57.8</td>
<td>62.7</td>
<td>55.7</td>
<td>49.5</td>
<td>52.4</td>
</tr>
<tr>
<td>Gold-standard</td>
<td>Head word</td>
<td>77.6</td>
<td>75.2</td>
<td>76.3</td>
<td>76.3</td>
<td>67.8</td>
<td>71.8</td>
</tr>
<tr>
<td></td>
<td>Boundary</td>
<td>74.4</td>
<td>66.5</td>
<td>70.2</td>
<td>67.5</td>
<td>60.0</td>
<td>63.5</td>
</tr>
</tbody>
</table>
Conclusion

• CCG helps find long-distance dependencies

• Performance on non-core arguments lower due to:
  – mismatches between CCGBank and PropBank annotation
  – sparser CCG feature set

Future Work:

• Use PropBank annotation in conversion to CCG