Language-Independent Parsing with Empty Elements

Shu Cai, David Chiang and Yoav Goldberg

6/21/2011
Trees without empty elements

```
(who \ is \ NNP \ VBN \ S
  |     |     |  |  |
  VP   NP   VP   S
  |     |     |  |  |
  WHNP VBZ NNP VBN S
  |     |     |  |  |
  WP   VBP NP VP
  |     |     |  |  |
  SBARQ
```
Trees with empty elements

```
SBARQ
  WHNP  SQ
    WP    VP
     who  is  NP  VBN  S
     VP    NP  VBN  VP
       John believed NP  VP
       -NONE-  TO  VB  -NONE-
           *  to  admirable  *T*
```
Trees with empty elements
Trees with empty elements
Empty Elements in English

**Trace:**
The apple which John ate (the apple)....

**Null Complementizer:**
The scientist said (that) the project ...

**Subject-control:**
John wants (John) to see...

And more..
Empty Elements in Chinese/Korean

- Trace and null complimentizer like English
- “Pro-drop” languages
Empty Elements in Chinese/Korean

- Trace and * like English
- “Pro-drop” languages
Empty Elements in Chinese/Korean

- Trace and * like English
- “Pro-drop” languages
Our Task

Parser which can deal with empty elements

The apple he ate
Our Method

Lattice Parsing  
(Chappelier et al. 1999)  
+  
Slash transformation  
(Schmid 2006)  
+  
State of the art parser  
(Petrov et al. 2006)
Training

(a)
Training
Training

(a) SBARQ
   WHNP
   |    VP
   |    VP
   |    NP
   |    to
   |    VP
   |    to
   WP    VBZ    NNP    VBN
   who    is    John    believed

(b) SBARQ
   WHNP
   |    VP
   |    VP
   |    NP
   |    to
   |    VP
   |    to
   WP    VBZ    NNP    VBN    S
   who    is    John    believed    NP

-NONE-

* 

T*
Training

Slash feature: (Schmid, 2006)
Testing

- Insert empty element symbol (*e*) to produce word lattice
- Parse over word lattices
Evaluation Measures

- Precision/Recall/F-1 of two multisets

- Labeled empty elements: \((t;i;i)\) for each empty element

  \textit{The apple he ate} *T* \(\rightarrow\) (*T*,4,4)

- Labeled brackets: \((X;i;j)\) for each nonterminal node (non-empty and empty)
Evaluation measures in previous work (for comparison)

English: Labeled empty brackets: \((X/t;i;i)\) for each empty nonterminal node (Johnson, 2002)

Chinese: Unlabeled empty elements: \((Yang \text{ and Xue, 2010})\)
Experiments

- Standard settings for English and Chinese parsing
- Try to be consistent with previous work
- Use multiple evaluation measures
## English (Test set)

<table>
<thead>
<tr>
<th></th>
<th>Labeled Empty Brackets</th>
<th>Labeled Empty Elements</th>
<th>All Labeled Brackets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P  R  F1</td>
<td>P  R  F1</td>
<td>P  R  F1</td>
</tr>
<tr>
<td>Schmid(2006)</td>
<td>86.1 81.7  <strong>83.8</strong></td>
<td>87.9 83.0  <strong>85.9</strong></td>
<td>86.8 85.9  <strong>86.4</strong></td>
</tr>
<tr>
<td>Our result</td>
<td>90.1 79.5  <strong>84.5</strong></td>
<td>92.3 80.9  <strong>86.2</strong></td>
<td>90.1 88.5  <strong>89.3</strong></td>
</tr>
</tbody>
</table>
## Chinese (Test set)

<table>
<thead>
<tr>
<th></th>
<th>Unlabeled Empty Elements</th>
<th>Labeled Empty Elements</th>
<th>All Labeled Brackets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P  R  F1</td>
<td>P  R  F1</td>
<td>P  R  F1</td>
</tr>
<tr>
<td>Yang and Xue(2010)</td>
<td>80.3 57.9 63.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Our result</td>
<td>74.0 61.3 67.0</td>
<td>66.0 54.5 58.6</td>
<td>82.7 80.8 81.7</td>
</tr>
</tbody>
</table>
Conclusion

- Empty elements is an integral part of syntax
- PCFG+Lattice representation provides a way to recover empty elements from plain text
- Improvement of empty element recovery accuracy on English/Chinese
Future Work

- Explore Trace & co-indexation information (the part is moving to where?)

- Language-dependent issues: will special treatment of *pro* and *PRO* make Chinese empty element parsing better?
Parser we use:
http://www.cs.bgu.ac.il/~yoavg/software/blatt/

Evaluation code:
http://www.isi.edu/~chiang/software/eevalb.py