Language-Independent Parsing with Empty Elements

Shu Cai, David Chiang and Yoav Goldberg

6/21/2011
Trees without empty elements

[Tree diagram showing the syntactic structure of a sentence]
Trees with empty elements

```

[Diagram of a tree with empty elements]

S
  /
SBARQ
  /
WHNP
   /
WP
    /
who

SQ
  /

VP
  /
NP
   /
John
    /
believed

S
  /
NP
  /
-NONE-
    /
TO
     /
to
      /

VP
  /
NP
   /
admire
    /
-NONE-
      /
*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
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

  \[\text{The apple he ate } *T* \rightarrow (*T*, 4, 4)\]

- 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 and Xue, 2010)
Experiments

- Standard settings for English and Chinese parsing
- Try to be consistent with previous work
- Use multiple evaluation measures
|                  | Labeled Empty Brackets | Labeled Empty Elements | All Labeled Brackets |
|------------------|-------------------------|------------------------|---------------------|
|                  | P  R  F1                | P  R  F1                | P  R  F1            |
| Schmid(2006)     | 86.1 81.7 83.8          | 87.9 83.0 85.9          | 86.8 85.9 86.4      |
| Our result       | 90.1 79.5 84.5          | 92.3 80.9 86.2          | 90.1 88.5 89.3      |
Chinese (Test set)

|                               | Unlabeled Empty Elements | Labeled Empty Elements | All Labeled Brackets |
|-------------------------------|--------------------------|------------------------|----------------------|
|                               | P  | R  | F1 | P  | R  | F1 | P  | R  | F1 |
| Yang and Xue (2010)           | 80.3 | 57.9 | 63.2 |       |     |    |     |     |     |
| Our result                    | 74.0 | 61.3 | 67.0 | 66.0 | 54.5 | 58.6 | 82.7 | 80.8 | 81.7 |
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