Engineering a Deep HPSG for Mandarin Chinese

Rui Wang
with Yi Zhang and Yu Chen
DFKI GmbH, Germany
A “Deep” Grammar

- Head-Driven Phrase Structure Grammar (HPSG)
  - [Pollard and Sag, 1994]
  - Unification-based formalism with typed feature structures
  - Detailed lexical type hierarchy
  - Generalized schemata (grammar rules)
    - head-subject, head-complement, head-adjunct, head-specifier, ...
  - Compositional semantics in MRS
Foundation

- Platform: LKB + TDL
- Based on Matrix customization
  - [Bender, 2008]
- Two reference books
  - [Zhu, 1986]
  - [Li and Thompson, 1989]
- Website: [http://mcg.opendfki.de/](http://mcg.opendfki.de/)
Head Types
(Interesting) Phenomena

- Topic construction
- Numeral-Classifier Phrases
- DE(的)-constructions
  - Associative DE-phrase
  - Nominalizing DE-phrase (including relatives)
- Locatives
- Simple BA-construction
- Simple BEI-construction
Topic construction

- Topic is analyzed as a second optional subject for predicatives
  - Subj2-head rule (from head-specifier rule)
  - Underspecified semantic role of the topic (future)
Numeral-Classifier Phrases

• Nouns have two specifiers
  \[ \text{SPR} < [\text{CLP}], [\text{DEM}] > \]

• Classifier takes numeral as specifier

• None of the specifiers is obligatory, but our rules define the order of them (if they occur)
Numeral-Classifer Phrases (cont.)

- Demonstrative specifier [DEM]
- Numeral classifier (phrase) [CL(P)]

```
那一朵花
that one CL flower
```
```
这个人
this CL person
```
```
(我吃了)个苹果
(I eat LE) CL apple
```
DE(的)-constructions

• Associative DE-phrase (Including possessive)
  ▫ Associate relation modifying (head-modifier)

• Nominalizing DE-phrase (including relatives)
  ▫ DE is the head of the DE-phrase, changing a predicative into a substantive, e.g. 红的(red thing)
DE(的)-constructions (cont.)

• Associative

```
我 (I) 的 (DE) 书 (book)
```

```
comp-h  mod-h
```

```
assoc(我, 书)
assoc(I, book)
```
DE(的)-constructions (cont.)

- Nominalizing
  - Subj gap
  - Obj gap
  - No gap
Coverb

- Coverb (as defined in Li & Thompson 1981)
  - Prepositions (introduce semantic relation)
    - e.g. ZAI
  - BA
  - BEI
Simple BA-construction

- **BA-fronted-phrase (unary)**
  - Move the right complement to the left side of the verb
  - Only applied to the right-most complement of the predicative, e.g. iobj cannot be fronted
Simple BEI-construction

- **BEI-passive-phrase (unary)**
  - Switch the semantic roles of the subj and obj (right-most)
Other Phenomena

- Resultative verb compound
- Aspect marking (着, 了, 过)
- Sentence final particle (question, 了)
- Basic serial verb constructions
- ......
## Evaluation

<table>
<thead>
<tr>
<th>Number of Sentences</th>
<th>129</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positives</td>
<td>110</td>
</tr>
<tr>
<td>Negatives</td>
<td>19</td>
</tr>
<tr>
<td>Average words / sentence</td>
<td>5.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gold Standard</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
</tr>
<tr>
<td>System</td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
</tr>
</tbody>
</table>
# Error Distribution

<table>
<thead>
<tr>
<th>Phenomena</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex BA &amp; BEI</td>
<td>11</td>
</tr>
<tr>
<td>Post-verbal nominative modifier</td>
<td>4</td>
</tr>
<tr>
<td>Verb-negation-Verb</td>
<td>4</td>
</tr>
<tr>
<td>Nominative predicate</td>
<td>3</td>
</tr>
<tr>
<td>Post-predicate adverbial modifier</td>
<td>2</td>
</tr>
<tr>
<td>Coordination</td>
<td>2</td>
</tr>
<tr>
<td>Adverbial suffix (地方)</td>
<td>1</td>
</tr>
<tr>
<td>Noun-noun compound</td>
<td>1</td>
</tr>
<tr>
<td>Others</td>
<td>1</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td><strong>28</strong></td>
</tr>
</tbody>
</table>
Development

- Developers: Yi Zhang, Rui Wang, Yu Chen
- About 200 newly added types
- Over 500 manually compiled lexical entries
- A (Chinese) phenomena-oriented testsuite with over 700 positive & negative items (only 129 included in the release)
- Released under the LGPL-LR (excluding the testsuite)
- Website: http://mcg.opendfki.de
Ongoing Work

• Lexicon acquisition
  ▫ From treebanks
  ▫ From existing dictionaries

• Test suites and Treebanks

• Broad-coverage parsing and generation
  ▫ CFG approximation (Zhang and Krieger, 2011)
References

- **Generic**
  - [Zhu, 1986], [Li and Thompson, 1989], [Bender, 2000]

- **HPSG**
  - [Ng, 1997], [Gao, 2000], [Müller and Lipenkova, 2009], [Yu et al., 2010]

- **LFG**
  - [Guo, 2007], [Guo, 2009]

- **CCG**
  - [Tse and Curran, 2010]
Thank YOU!

Questions?