CWN-Viz : Semantic Relation Visualization in Chinese Wordnet

*Ming-Wei Xu, **Jia-Fei Hong, 
***Shu-Kai Hsieh, *Chu-Ren Huang

* Institute of Linguistics Academia Sinica, Taiwan  
** Graduate Institute of Linguistics, National Taiwan University  
*** Department of English, National Taiwan Normal University
Outline

• Introduction
• Chinese Wordnet
• Applying visualization techniques to Lexical Semantic Relations
• Conclusion
Introduction

• WordNet (Miller et al, 1990)

• Sinica BOW (Huang et al, 2004)
  — The Academia Sinica Bilingual Ontological Wordnet (Sinica BOW)
  — It integrates three resources: WordNet1.6/1.7.1, English-Chinese Translation Equivalents Database (ECTED), and SUMO.

• Chinese Wordnet at Academia Sinica
Chinese Wordnet

- CWN Group (Huang, 2003–)
- Highly linguistically motivated criteria for sense distinction
  - Linguistic Felicity for sense/meaning facet
- Bilingual Ontological Lexical Resource
- LSRs bootstrapping from PWN/EWN
Data Structure of WN

- Lemma
- Sense
- Synset
- Semantic Relation
- Example
Data Structure of CWN

- **Lemma**
- **English Synset**
- **Sense**
- **Facet**
- **Domain**
- **Semantic Relation**
- **WordNet**
- **SUMO Concept**

- **Chinese Lemma**
- **Chinese Lexical Knowledge (Analyzed by CWN Group)**
- **Link to Related Language Resource**
Bootstrapping Lexical Semantic Relations

- Cross-lingual conversion of Lexical Semantic Relations via inference rules (Huang et al. 2005; Hsieh et al. 2006)

**Parallel Wordnets LSRs Conversion (complete model)**

Translation-mediated LSR: The unknown LSR $y = i + x + ii$

Diagram:
- CW2
- EW2
- CW1
- EW1

$\begin{align*}
  x &= EW1 - EW2 \\
  y &= CW1 - CW2 \\
  i &= \text{Translation LSR} \\
  ii &= \text{Translation LSR} \\
  \text{The unknown LSR:} \quad y &= i + x + ii
\end{align*}$
Parallel Wordnets LSRs Conversion (reduced model)

Translation-mediated LSR (When TEs are synonymous): The unknown LSR $y = 0 + x + 0 = x$ (for $i = 0$; $ii = 0$)
Example (When TEs are synonymous):

- Fat (00934421A) = 脂肪的 (feipang gde)
- Thin (00936334A) = 瘦 (shou)
- Chubby (00935062A) = 肥胖的 (feipang gde)
- CW1 = EW1 (i = 0)
Example (When TEs are not synonymous):

- talk (00638109V) = 說 (shou)
  
  - x = HYP (‘talk’ is the hypernym of ‘whisper’)
  
  - y
  
  - i = HYP (‘whisper’ is the hyponym of ‘er-yu’)
  
- whisper (00248910N)

- baklava (05674827N) = 果蜜千層餅 (guo-mi-qian-ceng-bing)
  
  - x = HPO (‘baklava’ is the hyponym of ‘pastry’)
  
  - y
  
  - i = HPO (‘pastry’ is the hyponym of ‘gao-dian’)

- pastry (05670938N)

- 印度 (er-yu)

- 耳語 (er-yu)

- 糕點 (gao-dian)
<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>X</th>
<th>Y</th>
<th>Bootstrapped Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HYP</td>
<td>ANT</td>
<td>ANT</td>
<td>{CW1, ANTONYM, CW2}</td>
</tr>
<tr>
<td>2</td>
<td>HYP</td>
<td>HYP</td>
<td>HYP</td>
<td>{CW1, HYPOXYMY, CW2}</td>
</tr>
<tr>
<td>3</td>
<td>HYP</td>
<td>NSYN</td>
<td>HYP</td>
<td>{CW1, HYPOXYMY, CW2}</td>
</tr>
<tr>
<td>4</td>
<td>HYP</td>
<td>HOL</td>
<td>HOL</td>
<td>{CW1, HOLONYMY, CW2}</td>
</tr>
<tr>
<td>5</td>
<td>HYP</td>
<td>all other LSRs</td>
<td>undecided</td>
<td>?</td>
</tr>
<tr>
<td>6</td>
<td>HPO</td>
<td>ANT</td>
<td>ANT</td>
<td>{CW1, ANTONYM, CW2}</td>
</tr>
<tr>
<td>7</td>
<td>HPO</td>
<td>HPO</td>
<td>HPO</td>
<td>{CW1, HYPOXYMY, CW2}</td>
</tr>
<tr>
<td>8</td>
<td>HPO</td>
<td>NSYN</td>
<td>HPO</td>
<td>{CW1, HYPOXYMY, CW2}</td>
</tr>
<tr>
<td>9</td>
<td>HPO</td>
<td>MER</td>
<td>MER</td>
<td>{CW1, MERONYMY, CW2}</td>
</tr>
<tr>
<td>10</td>
<td>HPO</td>
<td>all other LSRs</td>
<td>undecided</td>
<td>?</td>
</tr>
<tr>
<td>11</td>
<td>NSYN</td>
<td>ANT</td>
<td>ANT</td>
<td>{CW1, ANTONYM, CW2}</td>
</tr>
<tr>
<td>12</td>
<td>NSYN</td>
<td>HYP</td>
<td>HYP</td>
<td>{CW1, HYPERNYMY, CW2}</td>
</tr>
<tr>
<td>13</td>
<td>NSYN</td>
<td>HPO</td>
<td>HPO</td>
<td>{CW1, HYPOXYMY, CW2}</td>
</tr>
<tr>
<td>14</td>
<td>NSYN</td>
<td>NSYN</td>
<td>NSYN</td>
<td>{CW1, NEAR-SYNONYMY, CW2}</td>
</tr>
<tr>
<td>15</td>
<td>NSYN</td>
<td>MER</td>
<td>MER</td>
<td>{CW1, MERONYMY, CW2}</td>
</tr>
<tr>
<td>16</td>
<td>NSYN</td>
<td>HOL</td>
<td>HOL</td>
<td>{CW1, HOLONYMY, CW2}</td>
</tr>
<tr>
<td>17</td>
<td>HOL</td>
<td>ANT</td>
<td>ANT</td>
<td>{CW1, ANTONYM, CW2}</td>
</tr>
<tr>
<td>18</td>
<td>HOL</td>
<td>HYP</td>
<td>HYP</td>
<td>{CW1, HYPOXYMY, CW2}</td>
</tr>
<tr>
<td>19</td>
<td>HOL</td>
<td>NSYN</td>
<td>HOL</td>
<td>{CW1, HOLONYMY, CW2}</td>
</tr>
<tr>
<td>20</td>
<td>HOL</td>
<td>HOL</td>
<td>HOL</td>
<td>{CW1, HOLONYMY, CW2}</td>
</tr>
<tr>
<td>21</td>
<td>HOL</td>
<td>all other LSRs</td>
<td>undecided</td>
<td>?</td>
</tr>
<tr>
<td>22</td>
<td>MER</td>
<td>ANT</td>
<td>ANT</td>
<td>{CW1, ANTONYM, CW2}</td>
</tr>
<tr>
<td>23</td>
<td>MER</td>
<td>HPO</td>
<td>HPO</td>
<td>{CW1, HYPOXYMY, CW2}</td>
</tr>
<tr>
<td>24</td>
<td>MER</td>
<td>NSYN</td>
<td>MER</td>
<td>{CW1, MERONYMY, CW2}</td>
</tr>
<tr>
<td>25</td>
<td>MER</td>
<td>MER</td>
<td>MER</td>
<td>{CW1, MERONYMY, CW2}</td>
</tr>
<tr>
<td>26</td>
<td>MER</td>
<td>all other LSRs</td>
<td>undecided</td>
<td>?</td>
</tr>
</tbody>
</table>
Current States

- From 2003 till September 2007,
  - 7198 lemma
  - 17932 senses
  - 4191 mapping to PWN synset
    \( WN2.0:4134/WN2.1:30/WN3.0:27 \)
  - 13823 Synsets
  - 18006 Relations
Applying Visualization Technique to LSRs

- Previous works: WordNet TreeWalk (Bou, 2003); WordNet Connect (Fong, 2003); WordNet Relationship Browser (Alcock, 2004).

- More recently, Visual Thesaurus (ThinkMap, 2005); Visual WordNet Project (Kuo, 2005); WordNet Explorer (Collins, 2006)
However, we need alternative tools to meet our needs:

- Not only show a full picture of WordNet relations, but also give context as well. (e.g., corpus, dictionary gloss, thesaurus, etc.)
- Distinctively show predicted LSRs with via bootstrapping rules (Huang et al, 2003) for the purpose of evaluation.
- Provide a window for showing concept clusters using morpho-semantic links
CWN-Viz: The First Try

- Interface for browsing integrated resources and evaluating bootstrapped LSRs.

- Technically, we follow the design paradigm based on “TouchGraph” (Google) - an open source graph layout system, to construct a working prototype of a visualization suite for Chinese Wordnet.

- For now, it can show all lemmas, senses, and semantic relations for a word form recorded in Chinese Wordnet, and basic measure of the distance of each semantic relation.
The basic visualization construction
Visualization construction of Semantic relations
Calculating Principle

- A keyword root: a center node and extend two levels
- Based on the first level nodes to calculate sub-roots of the sub-trees
- These sub-trees
  - Evaluate the relationship score for each sub-tree
- Calculate the relationship score for each sub-tree
  - Present the calculating matrix for each cluster
- Select the most nodes of the numbers semantic relations until all nodes
The clusters --- Viz construction
The second lemma of 正
The sixth senses of 正
The first meaning facet of the first sense of 美麗
Viz for 正 Zheng4 ('right')
GWC-08, Szeged Hungary, Jan. 22–25, 2008
Conclusion

• Design the visualization tool such that various language resources and cross-lingual lexical semantic relations be processed.

• Facilitate linguists’ use and understanding of Wordnets.
Thank You!

- Sinica BOW:
  http://bow.sinica.edu.tw
- Chinese Wordnet
  http://cwn.ling.sinica.edu.tw
- CWN-Viz Prototype Demo
  http://cwn.ling.sinica.edu.tw/cwnviz/