GuanXi-Network

A new Multilingual LLOD

Applications for
Language Learning and NLP
Introduction

• Linguistic resources in Language Learning and NLP

• Challenges
  - Link scattered data
  - Words, Phrases, or Sentences
  - How to map meaning?

• Solution
  - Linguistic Linked Data technologies
Outline

1. Approach and Resources
2. Description of resources
3. Manual evaluations and mapping issues
4. Case study and discussion
Approach

• Models
  – Lemon
  – Translation.owl

• Tool
  – GuanXify (Ruby wrapper)
    • Input: word-sense pair
    • Output: lemon lexicons + translation lexicons
Integrated Resources

- English PDEV Pattern Dictionary (mono)
- Chinese COW WordNet (bilingual)
  - ZH-EN
- Chinese CEDICT Dictionary (bilingual)
  - ZH-EN
- Ukrainian Slovnyk Dictionary (bilingual)
  - UK-ES
  - UK-EN
3 possible linking methods

1. **BabelNet as Proxy: Babelfy disambiguation API**

2. Sense extension of BabelNet
   - Automatic creation of sense keys from translation pairs in bilingual dictionaries

3. Resources interlinking by mapping English lexicons
Method: BabelNet as Proxy

1. BabelNet as Proxy: Babelfy disambiguation API
   - Pattern Dictionary:
     • Input: Verbs + 5 examples per pattern
     • Output: best sense for the verb
   - Chinese Dictionaries
     • Word segmentation + pos tagging for Chinese phrases
   - Ukrainian-Spanish-English
     • Input: phrases
     • Output: all senses for the words in the phrase/whole phrase
English PDEV Dictionary

- Statistics:
  - 1,254 verbs, 4,530 patterns, 3.61 patterns per verb
  - 19,651 triples (verbs, patterns, examples)

- Method
  - For each verb 5 examples are disambiguated using Babelfy

<table>
<thead>
<tr>
<th>POS</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>261</td>
</tr>
<tr>
<td>a</td>
<td>337</td>
</tr>
<tr>
<td>n</td>
<td>1,114</td>
</tr>
<tr>
<td>r</td>
<td>3</td>
</tr>
<tr>
<td>v</td>
<td>17,936</td>
</tr>
</tbody>
</table>

#Total Entries 19,651
Wrong POS 0.09

<table>
<thead>
<tr>
<th>#Verbs</th>
<th>#Total Patterns</th>
<th>AVG Patterns per verb</th>
<th>#Total of the unique BabelNet senses</th>
<th>AVG BabelNet senses per verb</th>
<th>#Total of all possible BabelNet Senses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,25</td>
<td>4</td>
<td>4,435</td>
<td>3.54</td>
<td>2,652</td>
<td>2.11</td>
</tr>
</tbody>
</table>
# Ukrainian Slovnyk Dictionary

<table>
<thead>
<tr>
<th>Phrase Set</th>
<th>%</th>
<th># Total Phrases</th>
<th># Total Phrases with BabelNet sense</th>
<th># Total Phrases without BabelNet sense</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN PHRASES</td>
<td></td>
<td>8,776</td>
<td>8,297</td>
<td>479</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>94.54%</td>
<td>5.46%</td>
</tr>
<tr>
<td>ES PHRASES</td>
<td></td>
<td>84,365</td>
<td>67,015</td>
<td>17,350</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>79.43%</td>
<td>20.57%</td>
</tr>
<tr>
<td>UK PHRASES</td>
<td></td>
<td>11,201</td>
<td>8,241</td>
<td>2,960</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>73.57%</td>
<td>26.43%</td>
</tr>
</tbody>
</table>
Chinese CEDICT/COW Dictionary

- CEDICT

<table>
<thead>
<tr>
<th></th>
<th>EN</th>
<th>ZH</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN</td>
<td>10,978</td>
<td>9,556</td>
</tr>
</tbody>
</table>

- COW

<table>
<thead>
<tr>
<th></th>
<th>EN</th>
<th>ZH</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN</td>
<td>114,634</td>
<td>33,346</td>
</tr>
</tbody>
</table>
Chinese CEDICT Dictionary

- Number of Chinese-English mappings: **24,848**

- **36%** one side only
- **12%** two sides no overlap
- **11%** two sides overlapped
- **41%** no side
One Side Only

- 主办国 (n.)  →  host country (n.)

No babelnet Sense  → Babelfy

New words need to be added into babelnet.
No Side

- 主修 (v) to major in

No babelnet Sense

major(a.)

discard

Discard due to the different pos tags.

Babelfy
Two Sides without Overlap

- 侵入 (vv) 家宅 (nn) 者 (nn) \rightarrow housebreaker

Babelfy

invasion(bn:00047325n) \rightarrow \text{X} \rightarrow \text{housebreaker}(bn:00016618n)

Babelnet

Missing links due to different babel senses made by babelfy.
Manual evaluation of English mappings

• Verb *brand* (all results: sealang.net/lider)
  - 5 wrong, 1 unsure, 2 correct

• Examples for brand 4

[[Business_Enterprise]] packages [[Artifact = Commercial Product]] in a way that is easily recognizable

1. “One important method used to sell benefits is by branding products.”
   - **WN:** A name given to a product or service

2. “We don't just *brand* the cheapest tubes available, *brand* them in pairs and hang *them up* on a dealer's wall.”
   - **WN:** To accuse or condemn or openly or formally or *brand* as disgraceful
Case study
-- top new Chinese words recently borrowed by English
Similar concepts in other cultures [edit]

Sociologists have linked *guanxi* with the concept of *social capital* (it has been described as a *Gemeinschaft* value structure) political behavior.¹

- *Biat* in Russian culture
- *Wasta* in Middle Eastern culture
- *Socialismo* in Cuban culture
- *Old boy network* in Anglo-Saxon and Finnish culture
- *Dignitas* in ancient Roman culture
- *Ksharem* (literally ‘connections’) in Israeli culture. *Proteksia* (from the word ‘Protection’) is the use of *Ksharem* for person:
Chinese Dama refers to a group of Chinese middle-aged women who rushed to purchase gold as an investment in the year 2013 when the gold price plunged greatly, especially in April and October.
Another sense of Chinese *dama* -- *dama* “conquering” squares in the world
Text is not enough, image tells --

Tuhao, in Chinese "土豪", refers to the mainland Chinese who are rich but uncultured.
When Tuhao meets Dama w/o Guanxi?
Conclusions and Perspectives

- GuanXi-Network for LL and NLP
- Connect “Englishes”
- Solve mapping words with phrases, and their meanings
- Improve sense mapping
  - disambiguation techniques
  - extension with translation pairs as senses
- Include visual glosses (image, schema etc.) and etymology
Thanks GuanXi?