Cognitively Salient Relations for Multilingual Lexicography

Gerhard Kremer\textsuperscript{▲} Andrea Abel\textsuperscript{☆} Marco Baroni\textsuperscript{▲}

\textsuperscript{▲}CIMeC, University of Trento
\textsuperscript{☆}EURAC, Bolzano

August 24, 2008
Background

- Lexicography:
  words related to each other

- Computer use:
  electronic dictionaries

- Psychology:
  salient relations
  in semantic norms collection
Background

- Lexicography:
  words related to each other

- Computer use:
  electronic dictionaries

- Psychology:
  salient relations
  in semantic norms collection
Background

- Lexicography:
  words related to each other

- Computer use:
  electronic dictionaries

- Psychology:
  salient relations
  in semantic norms collection
**Project Aims**

Systematically extract

... cognitively most salient relations

... for concept classes (nouns, preliminary)

... based on the languages German and Italian

First step:

concept description elicitation experiment
Project Aims

Systematically extract
  ... cognitively most salient relations
  ... for concept classes (nouns, preliminary)
  ... based on the languages German and Italian

First step:
concept description elicitation experiment
Experiment Purpose

- Relation type distributions: stable across languages?
- Different relations for different concept classes?
  - Which relation types are most salient?
  - At which level of class granularity?

⇒ Basis for corpus-based extraction of concept-class-specific relation instances
Experiment Outline

Subjects: High school students from South Tyrol

Stimuli: 50 concepts (from 10 hierarchically organised classes)

Task: Describe concept to alien in short phrases, in 1 minute
Data Preparation

- Splitting of phrases
- Normalisation
- Transcription into English
- Mapping to relation types

Data Snippet

<table>
<thead>
<tr>
<th>German</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birne pear fruit eine Frucht</td>
<td>a_fruit</td>
</tr>
<tr>
<td>Birne pear fruit gelbgrün</td>
<td>is_yellow-green</td>
</tr>
<tr>
<td>Birne pear fruit ist weich</td>
<td>is_soft</td>
</tr>
<tr>
<td>Birne pear fruit hat Kerne im Inneren</td>
<td>has_seeds</td>
</tr>
</tbody>
</table>
Property Type Distributions

... Across Languages

G. Kremer, A. Abel, M. Baroni
Cognitively Salient Relations

7 / 13
Property Type Distributions

... Across Classes

German

Italian

G. Kremer, A. Abel, M. Baroni  Cognitively Salient Relations
Clustering

Unsupervised clustering (cluto tool):

- Best solution with 3-way clustering
- 3 super-classes:
  - mammals & birds,
  - fruit & vegetable,
  - all other man-made objects & body parts
- Exception: horse (misclustered in Italian)
Summary

- Similar type distributions between German and Italian
- Distinct patterns among concept classes
- Robust clustering into 3 super-classes
Conclusion

- Next: perception experiment
- Relation extraction from corpora
- Application: electronic dictionary ELDIT

# Classes and Concepts Used

<table>
<thead>
<tr>
<th>Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>mammal:</td>
<td>dog, horse, rabbit, bear, monkey</td>
</tr>
<tr>
<td>bird:</td>
<td>seagull, sparrow, woodpecker, owl, goose</td>
</tr>
<tr>
<td>fruit:</td>
<td>apple, orange, pear, pineapple, cherry</td>
</tr>
<tr>
<td>vegetable:</td>
<td>corn, onion, spinach, peas, potato</td>
</tr>
<tr>
<td>body part:</td>
<td>eye, finger, head, leg, hand</td>
</tr>
<tr>
<td>clothing:</td>
<td>chemise, jacket, sweater, shoes, socks</td>
</tr>
<tr>
<td>tool:</td>
<td>comb, broom, sword, paintbrush, tongs</td>
</tr>
<tr>
<td>vehicle:</td>
<td>bus, ship, airplane, train, truck</td>
</tr>
<tr>
<td>furniture:</td>
<td>table, bed, chair, closet, armchair</td>
</tr>
<tr>
<td>building:</td>
<td>garage, bridge, skyscraper, church, tower</td>
</tr>
</tbody>
</table>