

Cognitively Salient Relations for Multilingual Lexicography

Gerhard Kremer[▲] Andrea Abel[☼] Marco Baroni[▲]

[▲]CIMeC, University of Trento

[☼]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 [▶ see full list](#)

(from 10 hierarchically organised classes)

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

Data Preparation

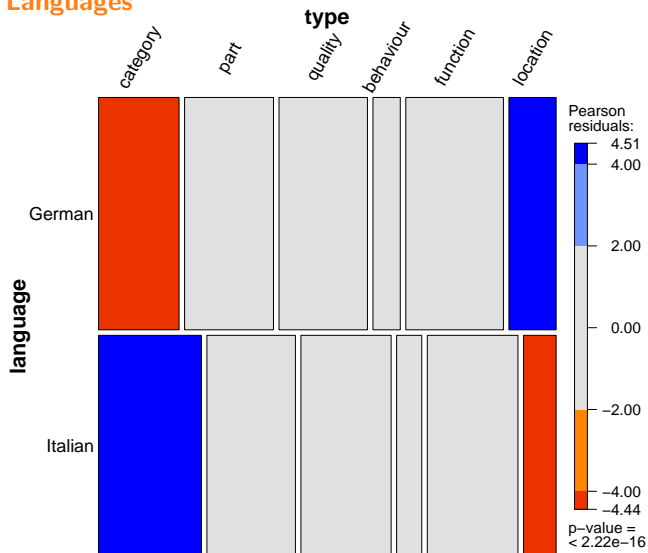
- Splitting of phrases
- Normalisation
- Transcription into English
- Mapping to relation types

Data Snippet

Birne	pear	fruit	eine Frucht	a_fruit	ch
Birne	pear	fruit	gelbgrün	is_yellow-green	ese
Birne	pear	fruit	ist weich	is_soft	ese
Birne	pear	fruit	hat Kerne im Inneren	has_seeds	eci

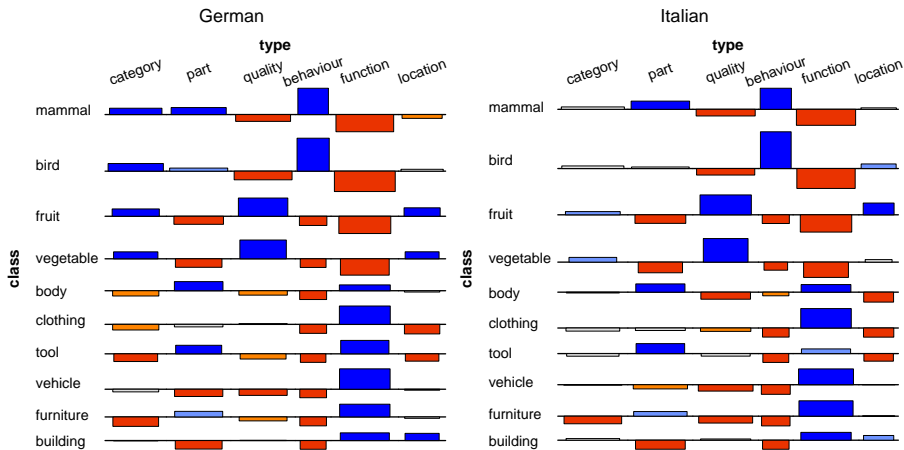
Property Type Distributions

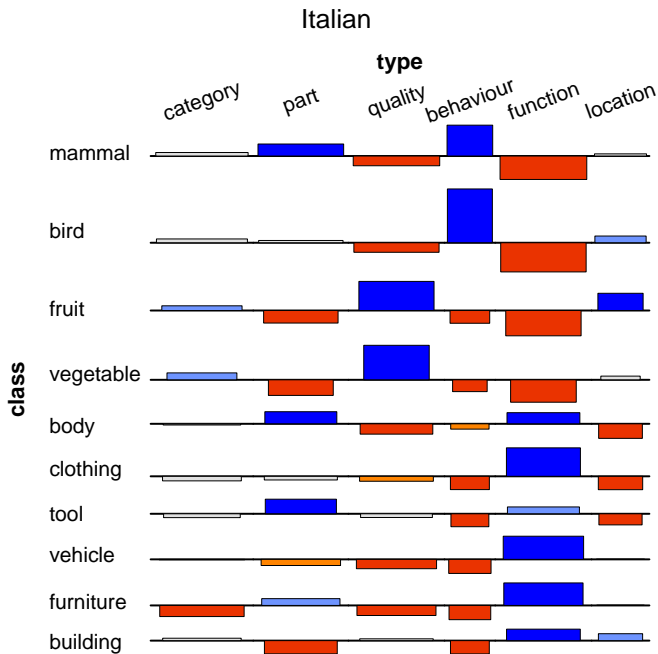
... Across Languages



Property Type Distributions

... Across Classes





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

Thank you

Bibliography



McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005).

Semantic Feature Production Norms for a Large Set of Living and Nonliving Things.

Behaviour Research Methods, Instruments & Computers, 37(4):547–559.



Vinson, D. P. and Vigliocco, G. (2008).

Semantic Feature Production Norms for a Large Set of Objects and Events.

Behaviour Research Methods, 40(1):183–190.

Classes and Concepts Used

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