Collecting Cognitively Salient Semantic Relations

Gerhard Kremer
gerhard.kremer@unitec.it

Introduction

Motivation
Electronic learner’s dictionary for German & Italian (ELDIT) – among others, shows semantically related words for a target word.

Examples
Concept dog:

- has paws, a tail, barks
- has a heart, breathes

Experiment: Feature Production

Task: Describe the given concept in short phrases
Stimuli: 50 concrete concepts from 10 classes (DE & IT)
Data: Categorised into types of relations (cf. McRae et al.)
Analysis: Deviations of the top 6 types from overall production frequency distribution (for each concept class)

Results
- Preferred relation types depending on which (super-) class the concept belongs to
- Similar patterns for German and Italian data

Goal
Semi-automatically harvest instances of cognitively salient semantic relations from text resources from the web.

Preliminary Focus

Composed part relations (adj_modifier + part-noun), e.g.
- rabbit: has long ears
- dog: has a wet nose

Method

Goal
Assuming that salient instances of part relations for a given concept have been identified already:
Find the most salient modifiers.

Production experiment data was used both as input (concept, part) and for evaluation (modifier-part pairs produced for a concept)

Approach
Based on occurrence frequencies of concept/part/modifier combinations in the WaCky Webcorpus, create ranked list and select the 5 highest ranked modifier candidates.

Webcorpus Excerpt Example

… Die mittelgroßen Affen leben in Gruppen von etwa 15 Tieren auf Bäumen im Regenwald. […] Die Kipunji sind verwandt mit anderen Mangaben, doch sie weisen einige Besonderheiten auf. Sie haben braunes oder helbraunes Fell und geben Töne von sich…

Evaluated Rank Lists (DE)

- modifiers [part] (within 4 words)
- within 20 sentences of concept
- same, but not considering concept context
- combination: multiplication of frequencies from both sources
- reranking: pull up those modifiers which are similar to those at higher ranks (by calculating their cosine distance based on nouns they co-occur with)

Production vs. Perception (DE)

Follow-up judgement experiment:
“The part of a concept is modifier.” – plausible or not?
Result for the acceptance rate of 0.75:
Small overlap of modifiers both produced and accepted (46), compared to modifiers only produced (53) or only accepted (42).

Conclusion

- Best method combines in-context and contextless information (with similar performance for German and Italian), and yields both produced and perceived modifiers
- Reranking improvable?
- Adaptable to part-noun collection and other relation types?

References


Advisors: Marco Baroni (CiMeC), Andrea Abel (EURAC)