

**A Crow's Beak is not Yellow –
Investigations on Cognitively Salient
Concept Properties**

Gerhard Kremer

January 18, 2011

Describing a Concept...



- ✓ has 2 legs
- ✓ has feathers
- ✓ is black

versus

- ⚡ has a heart
- ⚡ can see

Topic Outline

Feature norms (e. g. McRae et al.'s)

Concept representations – used in simulations of cognitive tasks

Efforts on extracting such descriptions

... using text corpora

(getting norms without experiments;

better models based on more data)

Topic Outline

Feature norms (e. g. McRae et al.'s)

Concept representations – used in simulations of cognitive tasks

Efforts on extracting such descriptions

... using text corpora

(getting norms without experiments;

better models based on more data)

~> Study aims

- Empirical basis: behavioural experiments (parallel: DE + IT)
- Method for harvesting cognitively salient properties

Experiments

Production

“Describe the concept.”

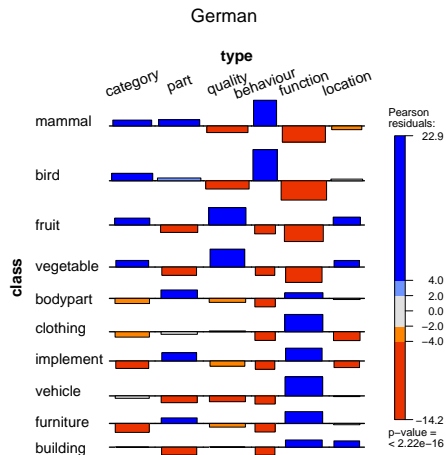
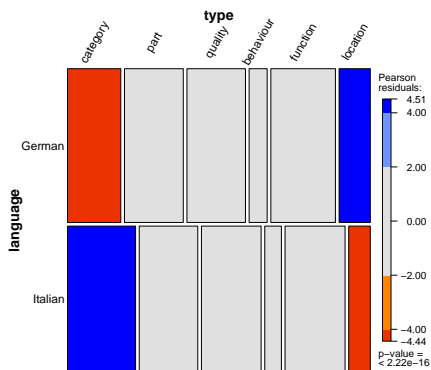
- Annotation of answers: property types
- Observation: preferred use of types depending on concept class

Perception

“Is the word plausible for describing the concept?”

- Recording of RTs and errors
- Findings: (inconsistent) differences between class-type pairs

Distributions of Produced Property Types



Experiments

Production

“Describe the concept.”

- Annotation of answers: property types
- Observation: preferred use of types depending on concept class

Perception

“Is the word plausible for describing the concept?”

- Recording of RTs and errors
- Findings: (inconsistent) differences between class-type pairs

Extraction Targets

Composite (adj-**modifier** + noun) **part** properties of **concepts**,

e. g. **rabbit**: **long** **ears**

Corpus excerpt

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 **hellbraunes** **Fell** und geben Töne von sich [...]

Approach

- **Aim:**

Extract cognitively salient modifiers
for given concept–part pairs

- **Idea:**

Create ranked list based on corpus frequencies
and select 5 highest ranked modifiers

- **Resource:**

WaCky web corpus

- **Evaluation** against feature production norms

Best Rank List Methods

1. Modifier–Part pair frequencies (“contextless”)

[*Adj*]? [*Adj*]? [*Adj*]? [*Adj*]? [*Noun*]

2. Frequencies of modifier–part pairs in concept context

[*part*]? (20 *sent.*) [*concept*] (20 *sent.*) [*part*]?

3. Productwise combination of frequencies

Example:

Concept “Bear” With Part “Fur”

rank	contextless		in concept context	
	freq	modifier	freq	modifier
1	507	thick	16	thick
2	209	dense	14	white
3	204	soft	11	small
4	185	black	11	soft
5	175	long	9	dense

Best Rank List Methods

1. Modifier–Part pair frequencies (“contextless”)

[*Adj*]? [*Adj*]? [*Adj*]? [*Adj*]? [*Noun*]

2. Frequencies of modifier–part pairs in concept context

[*part*]? (20 *sent.*) [*concept*] (20 *sent.*) [*part*]?

3. Productwise combination of frequencies

Best Rank List Methods

1. Modifier–Part pair frequencies (“contextless”)

[*Adj*]? [*Adj*]? [*Adj*]? [*Adj*]? [*Noun*]

2. Frequencies of modifier–part pairs in concept context

[*part*]? (20 *sent.*) [*concept*] (20 *sent.*) [*part*]?

3. Productwise combination of frequencies

Robust performance (precision $\approx 14\%$ at recall $\approx 43\%$)

- across languages,
- for production and perception,
- and for concepts previously unseen by the algorithm

Conclusion

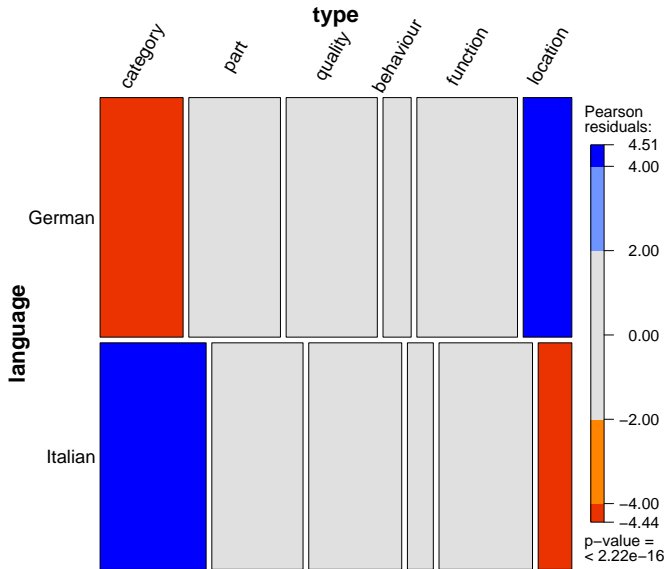
Types of concept properties produced by native speakers

- ... have similar distributions across languages
- ... are preferably from a type set depending on broad concept class

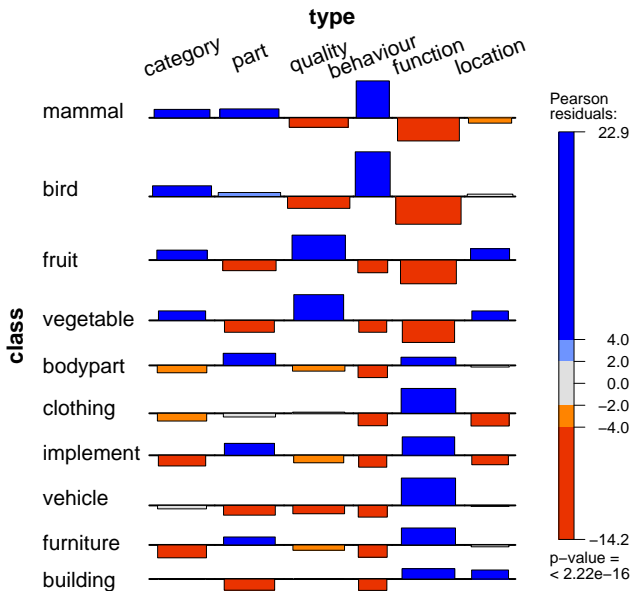
Automatic corpus-based extraction (of *part* modifiers)

- ... works best when combining in-context and contextless list
- ... performs similarly well across languages
- ... works comparably well based on both production and perception

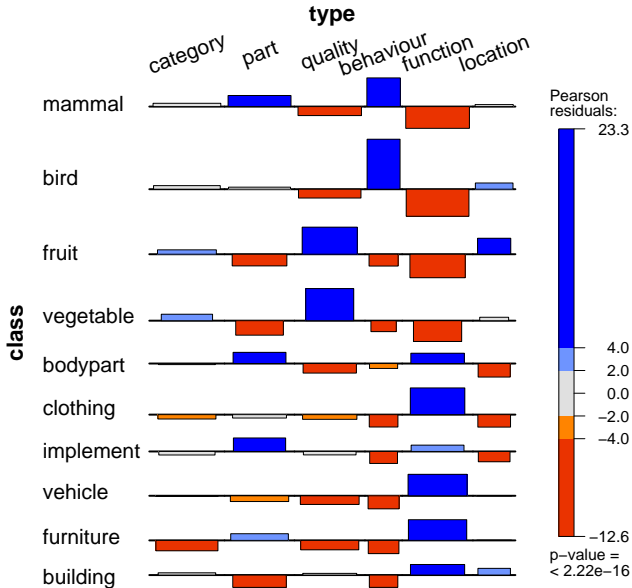
Thank you.



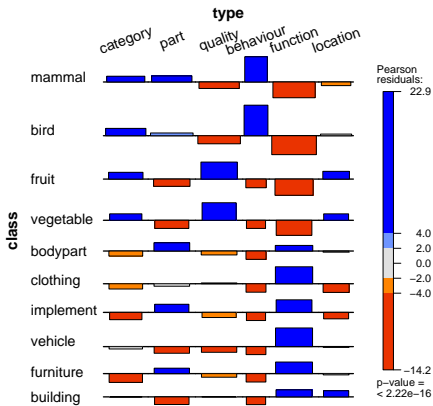
German



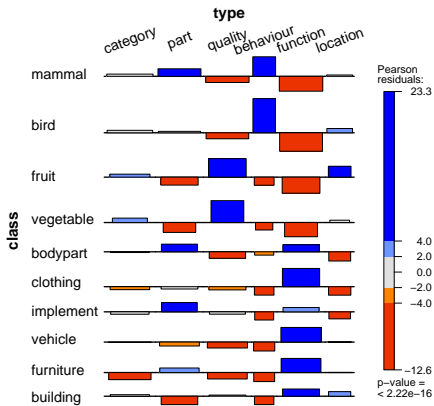
Italian



German



Italian



Algorithm Performances

