Defining the obvious: How to measure similarity?

Annika Tjuka
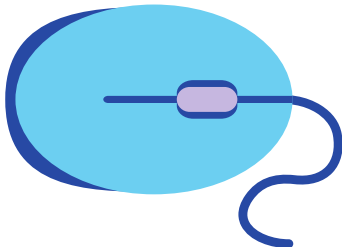http://annikatjuka.com
tjuka@shh.mpg.de

Max Planck Institute for the Science of Human History

September 26th, 2020

Polysemy

A polysemous word has multiple meanings:

mouse

Why?

Why do we use the same word for different meanings?

What cognitive processes lead to a word having multiple meanings?

Hypothesis

Polysemy is based on a perceived similarity between concepts.

The leg of the table

Similarity in linguistics

- **Conceptual metaphor theory**: similarity as the basis for metaphors (Lakoff and Johnson 1980)
- **WordNet**: similarity as the number of edges among sets of cognitive synonyms (synsets) (Fellbaum 1998)
- **Semantic neighborhoods**: similarity as the degree of alignment between words (Thompson et al. 2020)
- **CLICS**: similarity illustrated by colexification patterns across various languages (Rzymski et al. 2020)

Similarity in psychology

- **Feature** approach: similarity as an overlap between features (Tversky 1977)
- **Structure** approach: similarity as a constructive process in which two representations are structurally compared (Gentner and Markman 1997)
- **Transformation** approach: similarity as the number of steps that it takes to transform one mental representation into another (Hahn, Chater and Richardson 2003)

My research question

What determines the use of one word for different meanings across languages?

Method: Databases

Figure 1: https://concepticon.clld.org/
https://clics.clld.org/
http://digling.org/norare/
https://cldf.clld.org/

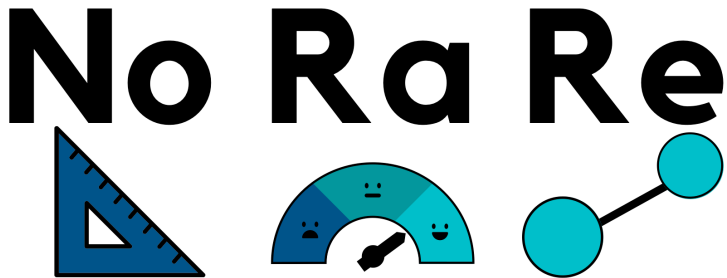## Step 1: Organize the available data

Figure 2: http://digling.org/norare/

Step 1: Organize the available data

Cross-Linguistic **No**rms, **Ra**tings, and **Re**lations
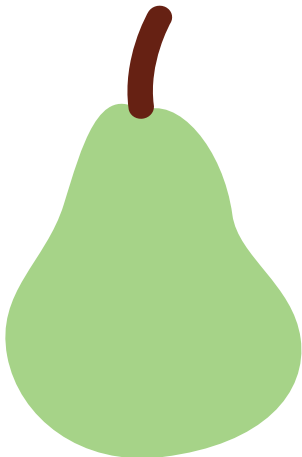for Words and Concepts (Tjuka, Forkel and List in preparation)

- **Norms**: e.g., word frequency, reaction time
- **Ratings**: e.g., age-of-acquisition, discrete emotions, sensory modality
- **Relation**: e.g., semantic field, polysemy

Step 1: Organize the available data

Statistics:

- approx. 40 new data sets psychology
- links to 3415 Concepticon concepts
- across 8 languages (i.e., English, Spanish, Dutch, Chinese)

Introduction
oooo

Similarity
oooo

Step 1
ooo

Step 2
●oo

Step 3
oooo

Open questions
o

Step 2: Define similarity measures

# Step 2: Define similarity measures

The Lancaster Sensorimotor Norms (Lynott et al. 2020):

- ratings on **perceptual modalities** (touch, hearing, smell, taste, vision) and five **action effectors** (mouth/throat, hand/arm, foot/leg, head excluding mouth/throat, torso)
- ratings are based on a 5-point scale

Step 2: Define similarity measures

Database of Cross-Linguistic Colexifications (CLICS, Rzymski et al. 2020):

- number of word colexifications across semantic categories
- colexification weight between concepts

Introduction
0000

Similarity
0000

Step 1
000

Step 2
000

Step 3
●000

Open questions
0
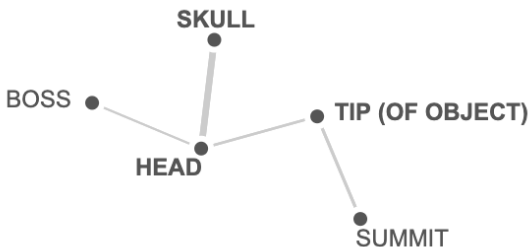
Step 3: Test hypothesis

Step 3: Test hypothesis

CLICS	Gloss	Dominant modality	Visual	Haptic	Hand
	HAND	Visual	4.25	3.65	4.42
300	ARM	Visual	4.58	4.16	4.75

Step 3: Test hypothesis

CLICS	Gloss	Dominant modality	Visual	Haptic	Head
	EYE	Visual	4.25	1.1	4.67
38	FACE	Visual	4.95	3.1	4.6

Step 3: Test hypothesis

CLICS	Gloss	Dominant modality	Haptic	Visual	Head
	AIR	Haptic	2.79	0.68	3.95
67	WEATHER	Visual	2.11	4.0	3.05
56	WIND	Haptic	3.69	1.06	3.53

Open questions

- Are the different definitions of similarity comparable?
- Is similarity measurable?
- Is similarity the only reason for polysemy?

References

- Christiane Fellbaum (1998) WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. American psychologist, 52(1), 45.

- Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. Cognition, 87(1), 1-32.

- Lakoff, G., & Johnson, M. (1980). Metaphors we live by. University of Chicago press.

- Rzymski, C., Tresoldi, T., Greenhill, S. J., Wu, M. S., Schweikhard, N. E., Koptjevskaja-Tamm, M., ... & Chang, S. (2020). The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies. Scientific data, 7(1), 1-12.

- Thompson, B., Roberts, S.G. & Lupyan, G. Cultural influences on word meanings revealed through large-scale semantic alignment.Nat Hum Behav (2020).

- Tjuka, A., Forkel, R., & List, J.-M. (in preparation). Linking Norms, Ratings, and Relations of Words and Concepts Across Multiple Language Varieties.

- Tversky, A. (1977). Features of similarity. Psychological review, 84(4), 327.