Similarity in linguistics and psychology

Annika Tjuka

Max Planck Institute for the Science of Human History

July 10th, 2020

Hypothesis

Polysemy is based on a perceived similarity between concepts.

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)

Similarity in linguistics

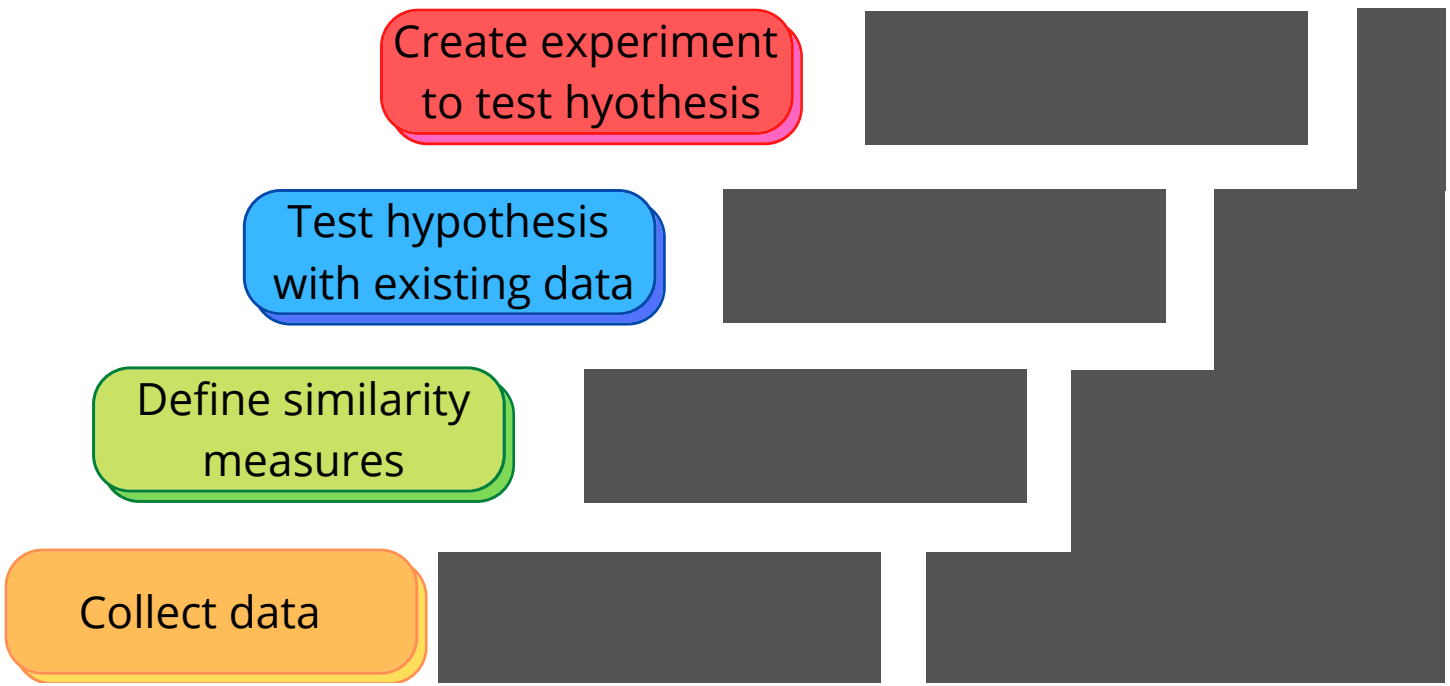
- Conceptual metaphor theory: similarity as the basis for metaphors (Lakoff and Johnson 1980)
- Semantic relatedness: similarity as the number of edges among WordNet synsets

The leg of the table

My research question

What determines the use of one word for different meanings?

Steps


Create experiment to test hyothesis
Test hypothesis with existing data
Define similarity measures
Collect data

Step 1: Collect data

Step 1: Collect data

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

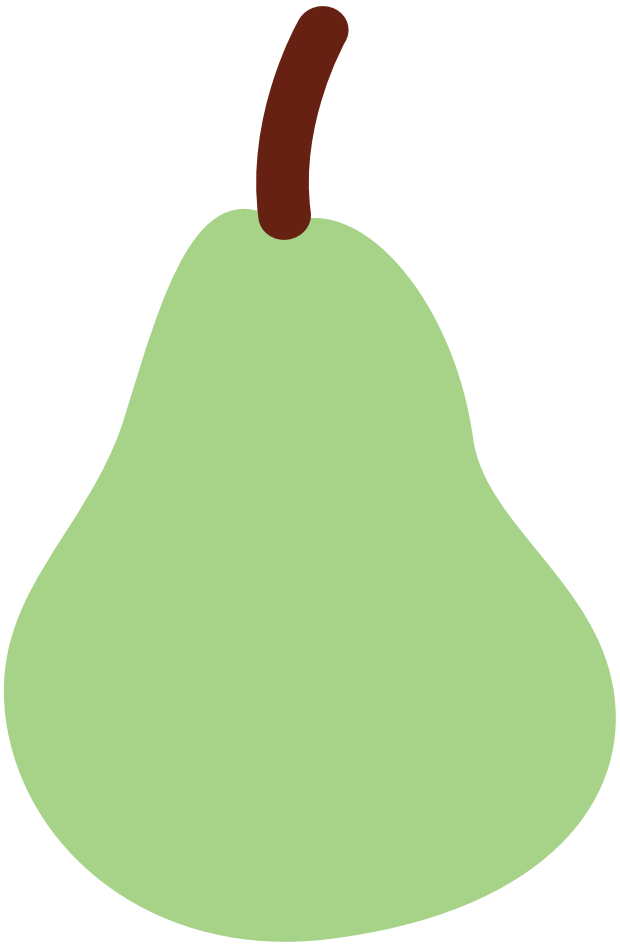
- **Norms**: e.g., word frequency, reaction time
- **Ratings**: e.g., age-of-acquisition, discrete emotions, sensory modality
- **Relation**: e.g., semantic field, polysemy

Step 1: Collect data

Statistics:

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

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 interoception) and five **action effectors** (mouth/throat, hand/arm, foot/leg, head excluding mouth/throat, and 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

Step 3: Test hypothesis

Step 3a: Test hypothesis within body part domain

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 3a: Test hypothesis within body part domain

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

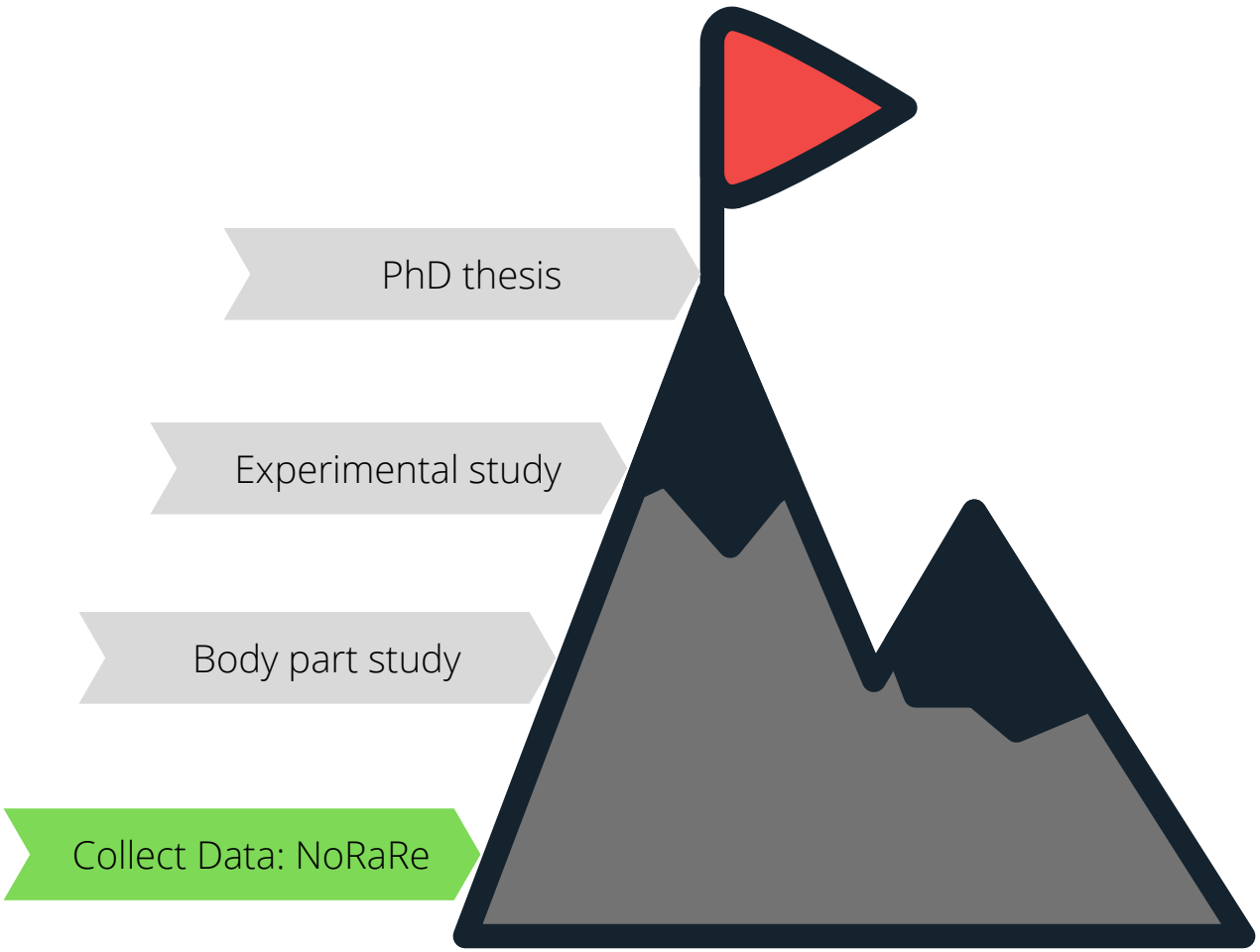
Step 3b: Test hypothesis in other domains

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

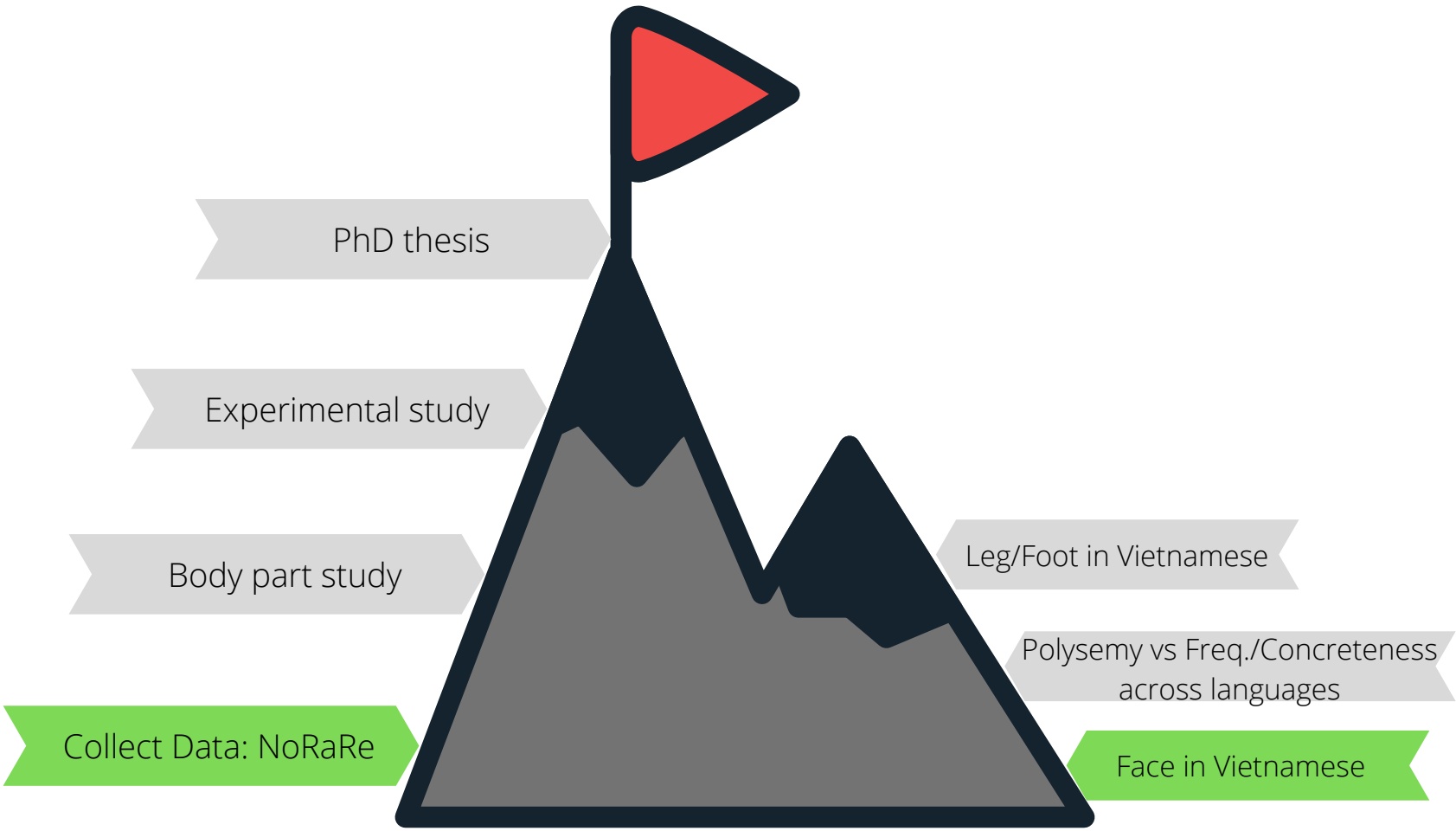
Step 3b: Test hypothesis in other domains

CLICS	Gloss	Dominant modality	Auditory	Head	Mouth
	ACCUSE	Auditory	3.65	2.91	2.95
64	BLAME	Auditory	3.25	2.29	2.24
30	CONDEMN	Auditory	3.6	2.65	2.45

Climbing to the top



Climbing the whole mountain



References

- 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.
- 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.