

A Concept List for the Study of Semantic Extensions from Body to Objects

Annika Tjuka

Department of Linguistic and Cultural Evolution

Max Planck Institute for the Science of Human History

Many words have multiple meanings, especially words for body parts. Some languages call the branch of a tree *arm* and others refer to the roof with the word *head*. These are instances of colexifications between body and object concepts and they occur across various languages. However, languages vary in terms of the frequency of body~object colexifications in their lexicon. It is, therefore, important to understand the different drivers that can lead to body~object colexifications. One possible factor to consider is the degree of how visually conspicuous a part of the body is. In an ongoing project, I investigate the role of visual salience as a driving force behind body~object colexifications. This blog post gives a brief overview and introduces the concept list that is the basis for my study.

Background

Studies in lexical semantic typology advance our understanding of the interplay between patterns that are dependent or independent of language structure. Thus, they allow insights into genealogical, cognitive, and cultural factors that lead to the different lexicons of the world's languages (for an overview, see Koptjevskaja-Tamm et al. [2007](#); Evans [2010](#)). Among the most frequently studied semantic domains are color, kinship, and the human body. Researchers are interested in how languages divide these different semantic domains into linguistic units, i.e., words. For example, languages show interesting differences in which parts of the body they refer to with a single word (Andersen 1978; Brown [1976](#); Enfield et al. [2006](#)). German has two words, *Hand* and *Arm*, whereas Vietnamese has a single word *tay* to refer to the concepts HAND and ARM. Thus, the Vietnamese language has a HAND~ARM colexification. Colexification occurs when two concepts are labeled with the same word and covers instances of polysemy and homonymy (François [2008](#)).

Interestingly, several studies that investigate colexifications in body terminology often discuss a few selected examples of colexifications between body and object concepts. Therefore, it seems to be the case that there are some common body~object colexifications occurring across languages, for example, eye~seed, testicle~egg (Brown & Witkowski [1981](#); [1983](#)). Presumably, all languages have at least a set of body~object colexifications in their vocabulary, but some languages, like Tzetal, use body~object colexifications productively and systematically to refer to different objects (Levinson [1994](#)). Colexifications between body and object concepts also occur in the languages of the Austronesian language family (Blust 2013) and some examples for Swahili and Polish are described in Kraska-Szlenk ([2014](#)). Based on an elicitation study across 13 languages, Tjuka (2019) showed that the languages Wolof and Mandarin Chinese have more body~object colexifications than Japanese and Marathi. A few body~object colexifications are also found in particular areas (Gast & Koptjevskaja-Tamm [2019](#)) and they play a role in language change (Urban [2011](#)).

One aspect mentioned as a determining factor for body~object colexifications across languages in the above literature is **visual salience**. The examples collected and analyzed in the studies often include outer body parts that are perceived visually. The features of those body parts, for instance, the long shape of the leg or the round shape of the head are assumed to increase the semantic extension of body part terms to objects. However, the following questions still remain unresolved:

1. Are visibly salient body parts more frequently colexified with objects across languages?
2. Are there differences in the frequencies and distribution patterns of certain body~object colexifications?
3. Do languages have a tendency to use more or less body~object colexifications?

In an ongoing project, I aim to answer these questions with a study on body~object colexifications. The colexifications were automatically computed for more than 1000 language varieties based on the *ClicsCore* subset of [Lexibank](#) (List et al. [2022](#)). The *ClicsCore* subset includes the data from CLICS (Rzymiski et al. [2020](#)) and 20 additional data sets. The basis for the study is a list of body and object concepts that were selected from the concept sets available in [Concepticon](#) (List et al. [2016](#); [2021](#)).

The Concept List

The concept list consists of 784 concepts divided into two groups: 134 body and 650 object concepts. The list was published on Zenodo and is available here: <https://doi.org/10.5281/zenodo.6365495> (see also Supplementary Material).

The body concepts are based on Tjuka (2020), but concepts referring exclusively to animal body parts such as [2965 PAW](#) or [72 CLAW](#) were excluded. Thus, the 134 body concepts were tagged as human body part in the SEMANTIC FIELD column. In addition, I included new body part terms that were added to [Concepticon](#) in a later version. The list is based on the concepts available in Version 2.5.0 of [Concepticon](#) (List et al. 2021).

The object concepts were further classified into ten semantic fields: animal, clothing, food, household items, instrument, landscape, plant, spatial relation, tool, and vehicle. Each semantic field category comprised a different number of concepts ranging from 169 concepts in the category tool (e.g., [945 SPEAR](#), [1074 FISHHOOK](#)) to 9 concepts in the category animal (e.g., [1065 SPIDER WEB](#), [3857 EGG SHELL](#)). Table 1 provides an overview of the total number of concepts in each semantic field category. Note that the animal category only included objects produced by animals, not animal body parts.

Table 1. Overview of the total number of concepts for each semantic field category and the number of visually salient versus non-visual concepts in each category.

| SEMANTIC FIELD | TOTAL | VISUAL | NON-VISUAL |
|------------------|-------|--------|------------|
| tool | 169 | 167 | 2 |
| human body part | 134 | 83 | 51 |
| landscape | 106 | 99 | 7 |
| food | 95 | 95 | 0 |
| household items | 95 | 95 | 0 |
| clothing | 64 | 64 | 0 |
| plant | 50 | 47 | 3 |
| vehicle | 28 | 25 | 3 |
| spatial relation | 19 | 15 | 4 |
| instrument | 15 | 15 | 0 |
| animal | 9 | 9 | 0 |

As shown in Table 1, the concepts were also categorized into two additional groups indicating whether they are visual or non-visual concepts. The human body part category included 83 visually salient concepts (e.g., [478 LIP](#), [1303 FINGER](#)) and 51 non-visual concepts (e.g., [1291 BACK](#), [3372 GALL BLADDER](#)). In the landscape category, 99 concepts were classified as visually salient (e.g., [3463 RIVERBANK](#), [3810 BOULDER](#)) and 7 as non-visual (e.g., [1667 HOLE](#), [3691 NEW MOON](#)). The spatial relation category comprised 15 visually salient (e.g., [863 EDGE](#), [1753 TOP](#)) and 4 non-visual concepts (e.g., [690 BOTTOM](#), [2323 BACK \(PART\)](#)). The semantic fields plant and vehicle included 3 non-visual concepts each. The two non-visual concepts in the tool category were [280 RIDGEPOLE](#) and [829 RAFTER](#). In the semantic fields food, household items, clothing, instrument, and animal, only visually salient concepts occurred. The overview in Table 1 illustrates that the concepts were unequally distributed across the visual versus non-visual category with a bias towards visually salient concepts.

Summary

Cross-linguistic comparison of body~object colexifications sheds light on genealogical, cognitive, and cultural factors that influence the semantic systems of the languages of the world. Visual salience plays an important role in meaning extensions and may be used across all languages. However, the variation in the frequency of body~object colexifications challenges the assumption of particular body~object colexifications such as EYE~SEED as being universal. The current project is a step towards the systematic study of body-object colexifications and new discoveries. The list presented in this blog post is used to study the frequency of colexifications across a wide range of languages.

Supplementary Material

Tjuka, Annika. 2022. A list of body and object concepts. Zenodo: <https://doi.org/10.5281/zenodo.6365495>.

References

- Andersen, Elaine S. 1978. Lexical universals of body-part terminology. In Joseph H. Greenberg (ed.), *Universals of Human Language: Word Structure*, vol. 3, 333–368. Stanford University Press.
- Blust, Robert. 2013. Body part terms and their extensions. In Robert Blust (ed.), *The Austronesian languages*. SEAsian Mainland Languages E-Series (SEAMLES), 321–328. Asia-Pacific Linguistics.
- Brown, Cecil H. 1976. General principles of human anatomical paronymy and speculations on the growth of paronymic nomenclature. *American Ethnologist* 3(3). 400–424. <https://doi.org/10.1525/ae.1976.3.3.02a00020>.

- Brown, Cecil H. & Stanley R. Witkowski. 1981. Figurative language in a universalist perspective. *American Ethnologist* 8(3). 596–615. <https://doi.org/10.1525/ae.1981.8.3.02a00110>.
- Brown, Cecil H. & Stanley R. Witkowski. 1983. Polysemy, lexical change and cultural importance. *Man* 18(1). 72–89. <https://doi.org/10.2307/2801765>.
- Enfield, Nicholas J., Asifa Majid & Miriam van Staden. 2006. Cross-linguistic categorisation of the body: Introduction. *Language Sciences* 28(2). 137–147. <https://doi.org/10.1016/j.langsci.2005.11.001>.
- Evans, Nicholas. 2010. Semantic typology. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 504–533. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199281251.001.0001>.
- François, Alexandre. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In Martine Vanhove (ed.), *From polysemy to semantic change: Towards a typology of lexical semantic associations* (Studies in Language Companion Series), vol. 106, 163–215. Amsterdam/Philadelphia: John Benjamins Publishing. <https://doi.org/10.1075/slcs.106.09fra>.
- Gast, Volker & Maria Koptjevskaja-Tamm. 2019. The areal factor in lexical typology. In Daniel Van Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of Linguistic Variation*, 43–82. Walter de Gruyter. <https://doi.org/10.1515/9783110607963-003>.
- Koptjevskaja-Tamm, Maria, Martine Vanhove & Peter Koch. 2007. Typological approaches to lexical semantics. *Linguistic Typology* 11(1). 159–185. <https://doi.org/10.1515/LINGTY.2007.013>.
- Kraska-Szlenk, Iwona. 2014. Semantic extensions of body part terms: Common patterns and their interpretation. *Language Sciences* 44. 15–39. <https://doi.org/10.1016/j.langsci.2014.02.002>.
- Levinson, Stephen C. 1994. Vision, shape, and linguistic description: Tzeltal body-part terminology and object description. *Linguistics* 32(4-5). 791–855. <https://doi.org/10.1515/ling.1994.32.4-5.791>.
- List, Johann-Mattis, Robert Forkel, Simon J. Greenhill, Christoph Rzymiski, Johannes Englisch & Russell D. Gray. 2022. Lexibank: A public repository of standardized wordlists with computed phonological and lexical features. *Research Square* (Preprint). <https://doi.org/10.21203/rs.3.rs-870835/v1>.
- List, Johann Mattis, Christoph Rzymiski, Simon Greenhill, Nathanael Schweikhard, Kristina Pianykh, Annika Tjuka, Carolin Hundt & Robert Forkel. 2021. *Concepticon. A resource for the linking of concept lists* (Version 2.5.0). Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology. <https://doi.org/10.5281/zenodo.596412>. <https://concepticon.clld.org/>.
- List, Johann-Mattis, Michael Cysouw & Robert Forkel. 2016. Concepticon: A resource for the linking of concept lists. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odiijk & Stelios Piperidis (eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, 2393–2400. Portorož, Slovenia: European Language Resources Association. <https://aclanthology.org/L16-1379/>.
- Rzymiski, Christoph, Tiago Tresoldi, Simon J. Greenhill, Mei-Shin Wu, Nathanael E. Schweikhard, Maria Koptjevskaja-Tamm, Volker Gast, Timotheus A. Bodt, Abbie Hantgan, Gereon A. Kaiping, Sophie Chang, Yunfan Lai, Natalia Morozova, Heini Arjava, Nataliia Hubler, Ezequiel Koile, Steve Pepper, Mariann Proos, Briana Van Epps, Ingrid Blanco, Carolin Hundt, Sergei Monakhov, Kristina Pianykh, Sallona Ramesh, Russell D. Gray, Robert Forkel & Johann-Mattis List. 2020. The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies. *Scientific Data* 7(1). 1–12. <https://doi.org/10.1038/s41597-019-0341-x>.

Tjuka, Annika. 2019. *Body-part metaphors as a window to cognition: A cross-linguistic study of object and landscape terms*. Humboldt-Universität zu Berlin, Master's thesis. <https://doi.org/10.17613/j95n-c998>.

Urban, Matthias. 2011. Asymmetries in overt marking and directionality in semantic change. *Journal of Historical Linguistics* 1(1). 3–47. <https://doi.org/10.1075/jhl.1.1.02urb>.