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 (Rzymski 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 EGGSHELL). 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.

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


References


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](https://doi.org/10.1075/jhl.1.1.02urb).