Annika Tjuka* and Johann-Mattis List

Partial colexifications reveal directional tendencies in object naming

https://doi.org/10.1515/gcla-2024-0005

Abstract: Expressions in which the word for a body part is also used for objects can be found in many languages. Some languages use body part terms to refer to object parts, while others have only a few idiosyncratic examples in their vocabulary. Studying the word forms referring to body and object concepts, i.e., colexifications, across languages, offers insights into cognitive principles facilitating such usage. Previous studies focused on full colexifications in which the same word form expresses two distinct concepts. Here, we utilize a new approach that allows us to analyze partial colexifications in which a concept is built out of the word forms for two separate concepts, like *river mouth*. Based on a large lexical database, we identified body and object concepts and analyzed 39 colexifications across 329 languages. The results show that word forms for body concepts are used slightly more frequently as a source for object names. However, the detailed examination of directional tendencies and colexifications of word forms between body and object concepts reveals linguistic variation. The study sheds light on meaning extensions between two concrete domains and showcases the synergies that arise through the combination of existing data and methods.

Keywords: Lexical typology, Body parts, Colexifications, Embodiment, Directionality

1 Introduction

Body part terms often have polysemous meanings (Dingemanse 2009; Kraska-Szlenk 2014). Apart from emotions and other abstract semantic domains, one area in which body part terms are used is in the naming of objects. English has multiple examples where a body part term refers to an object part, for example, *river*

* **Corresponding author: Annika Tjuka,** Max Planck Institute for Evolutionary Anthropology, Department of Linguistic and Cultural Evolution, Leipzig, Germany, annika_tjuka@eva.mpg.de
Johann-Mattis List, Max Planck Institute for Evolutionary Anthropology, Department of Linguistic and Cultural Evolution, Leipzig, Germany, mattis.list@uni-passau.de / University of Passau, Chair for Multilingual Computational Linguistics, Passau, Germany, mattis.list@uni-passau.de

mouth, *clock face*, or *foot of the mountain*. However, there seems to be no systematic pattern. In contrast, Tzeltal, a Mayan language spoken in Mexico, uses body part terms systematically based on a geometrical algorithm to refer to object parts (Levinson 1994). A recent study by Tjuka (2024b) showed that many languages have words that express body and object concepts in their vocabulary. The use of one word form for two distinct concepts across languages is called "colexifications" (François 2008). Tjuka (2024b) examined 78 body-object colexifications across 396 languages and demonstrated that visual similarity leads to widespread body-object colexification patterns. Apart from this general tendency, there was great linguistic variation and different dimensions of similarity led to language-specific colexification patterns. One limitation of the study was that it only included full colexifications, for example, Burmese *khoṅ* 'head, roof' and not partial colexifications such as Zapotec *gɨk* 'head' and *yo͡ʔo* 'house' which form a compound meaning ROOF. Here, we present a follow-up study investigating 39 partial colexifications of body and object concepts and the directionality of the meaning extensions across languages.

The words for parts of the human body provide an essential semantic domain for cognitive linguistics. However, most of the literature focuses on individual languages and the meaning extension of body part terms to abstract domains, for example, emotions (Baranyiné Kóczy and Sipőcz 2023; Pattillo and Waśniewska 2023; Baş and Kraska-Szlenk 2022; Brenzinger and Kraska-Szlenk 2014; Kraska-Szlenk 2020; Yu 2002; Ponsonnet 2014; Ponsonnet and Laginha 2020). Only a few studies describe non-figurative uses of body part terms for objects (e.g., Ibarretxe-Antuñano 2012; Kraska-Szlenk 2014). The reason for the neglect of studies on the meaning extensions of body part terms to the concrete domain of objects stems from the explicit exclusion of expressions like *foot of the mountain* or *leg of the table* from the analysis as conceptual metaphors by Lakoff and Johnson (2003/1980, 54). Their claim that these expressions are not systematic in language and thought is based on an English-centric perspective. Other languages show great systematicity in the extension of meanings of body part terms to objects (Levinson 1994; Tjuka 2024b). These findings support the importance of typological studies in cognitive linguistics and emphasize the relevance of a cross-linguistic approach to identifying constraints on linguistic variation (Croft 2016). Here, we present a study that explores words expressing body and object concepts to define the source and target domain of these colexifications across languages.

1.1 Directionality in meaning extensions

According to Lakoff and Johnson (1999, 18), "[t]he same neural and cognitive mechanisms that allow us to perceive and move around also create our conceptual systems

and modes of reason." On the basis of this embodiment hypothesis, the body is the source of our mental representations. Therefore, the direction of meaning extension between the semantic domains of the human body and objects should proceed from the body part term to the object term. However, only a few studies in cognitive linguistics exist that have investigated the extension of meanings between two concrete domains (e.g., Winter and Srinivasan 2022). In comparison, historical linguists have documented many examples of meaning extensions between two concrete domains over time and across diverse languages (e.g., Wilkins 1996; Brown and Witkowski 1981; Urban 2011).

Semantic change refers to the change of the meaning of a word. The processes underlying this phenomenon are interpreted as being based on cognitive mechanisms. For example, the perception of contiguity between the concept HOOD expressed in the Latin word *capitium* is the basis for the development of the Spanish word *cabeza* for HEAD (Steinberg 2014, 258). Studies by Traugott and Dasher (2001) and Wilkins (1996) indicate that semantic changes are more regular than previously assumed. Wilkins (1996, 272) documented many examples of semantic changes between the domains of the human body and objects. Apart from body part terms being the sources, these examples include semantic changes in which the direction of a semantic change goes from the object domain to the human body domain. This includes metonymic semantic changes of words for clothing items in Dravidian languages: FOOTLING → FOOT, EARRING → EAR, and PUBIC TASSEL → PENIS. Similarly, Brown and Witkowski (1981, 603-604) showed that words expressing the concept EGG are extended to include the concept TESTICLE due to the similarity of the round shape. Furthermore, Koch (2008, 128) demonstrated that the words for EGG or BALL frequently extend their meanings to EYEBALL. These examples are in contrast to the embodiment hypothesis. However, no systematic study investigating the directionality of meaning extensions between the two concrete domains of the human body and objects exists to date.

A systematic investigation of directionality consists of two parts. First, the source and target domains need to be determined. Second, the causes of the meaning extensions need to be analyzed. Since both parts of the analysis are laborious, studies have so far concentrated on only one of the two parts. Urban (2011) examined a selection of pairs of concepts, including human body parts and objects such as SKIN-BARK, MOUTH-ESTUARY, and TONGUE-FLAME. Based on an analysis of the asymmetry in the morphological complexity of word forms for each of the concepts across 149 languages, Urban (2011) determined the target and source concept of semantic changes. For example, many languages have a morphologically complex word form for BARK (e.g., Yuki ʔ *ol šil* 'tree skin'), whereas only a few languages have a morphologically complex word form for SKIN including the word for BARK Urban (2011, 8). Based on this observation, Urban (2011) concluded that SKIN is the source

concept and BARK is the target concept. Winter and Srinivasan (2022) used Urban's judgments of source and target concepts and investigated the factors that lead to meaning extensions between two domains. They tested the claim that the direction of meaning extension is from concrete to abstract domains by comparing English concreteness ratings and word frequency norms. Their results showed that word frequency is a better predictor of semantic change than concreteness. Based on a large-scale historical corpus, Xu et al. (2017) showed that asymmetry in other scales between two concepts leads to meaning extensions from embodied to disembodied, external to internal, less valenced to more valenced in English. The studies suggest that it is likely to find more meaning extensions from the domain of the human body to the domain of objects and that multiple factors play a role in the emergence of semantic changes.

1.2 Partial colexifications

With the advancement of computer-assisted methods, the comparison of vocabularies across different languages has become more efficient. Especially the theoretical construct "colexification", introduced by François (2008), has attracted the interest of cognitive scientists and led to studies investigating the variation of words expressing distinct concepts with large-scale datasets including many diverse languages (e.g., Jackson et al. 2019; Xu et al. 2020; Brochhagen and Boleda 2022). The basis of these studies is the implementation of colexification networks, which show the connections between concepts in a semantic space. A prominent implementation of this approach is the Database of Cross-Linguistic Colexifications (CLICS, Mayer et al. 2014; List et al. 2018; Rzymski et al. 2020). CLICS includes methods and tools for data representation and analysis of colexifications across 3,156 language varieties. It was the first use case to demonstrate the benefits of curating datasets in a standardized format, i.e., the Cross-Linguistic Data Formats (CLDF, Forkel et al. 2018). Building on this achievement, List et al. (2022) presented a large collection of multilingual word lists: Lexibank. This resource was used to identify 78 body-object colexifications in 396 language varieties by Tjuka (2024b). These colexification networks were based on one-to-one matches where a single word form was colexified with two concepts.

The study by List (2023) presented the first methodical approach for automatically computing partial colexifications. Partial colexifications occur when a word consists of more than one morpheme, one of which expresses a different concept, for example, *river mouth*. It is not straightforward to infer partial colexifications from word lists because shared morphemes between words may reflect grammatical

distinctions, resulting in a noisy network with many coincidental colexifications. The methods and workflows provided by List (2023) allow researchers to extract different types of partial colexification patterns from multilingual word lists. The most relevant type of partial colexifications for the present study are *affix colexifications*, which occur when one word form denoting a certain concept recurs at the beginning or the end of another word form expressing another concept. Note that the term "affix colexification" corresponds to the definition of "affix" in computer science, not linguistics, since the method does not identify morphemes but merely detects formal similarities. The methods presented by List (2023) infer affix colexifications based on a common substring between two sequences. For example, the substring ABC is a common substring of XYZABC. If two sequences overlap in at least one common substring of length 3, the algorithm defines it as an affix colexification. In a recent study, Bocklage et al. (2024) used affix colexifications to test the predictions of directionality in semantic change made by Urban (2011) on a larger language sample and found support for the less strict version of his hypothesis. The study showed that overt marking as reflected in automatically inferred affix colexifications seems to predict semantic change to some degree, provided that full colexifications are also attested for the concept pairs in question. Bocklage et al. (2024) also included a comparison with data from the Database of Semantic Shifts (Zalizniak et al. 2024), which contains a collection of semantic shifts documented in the world's languages.

The underlying assumption of the present study is that directional tendencies can be predicted by analyzing cross-linguistic patterns of affix colexifications. This assumption is based on the hypothesis that the mechanisms driving semantic motivation in word formation exhibit similar directional preferences as those governing semantic change (Koch and Marzo 2007). Based on the results by Bocklage et al. (2024), we assume that there is a positive correlation between semantic motivation and semantic change and that we can infer the directions in a colexification network with the methods described in List (2023). The present study uses the methods and data presented in List (2023) and Tjuka (2024b) to investigate the directionality of partial colexifications between the semantic domains of the human body and objects. We test the prediction that meaning extensions move predominantly from the body domain to the object domain.

2 Materials and methods

We used the 100 body-object colexifications from Tjuka (2024b) as a seed list. The list was derived from an automated identification of full colexifications between

134 body concepts and 650 object concepts in 36 Lexibank datasets, covering 931 language varieties. Established computational methods in Lexibank (List et al. 2022) were used for the study and the output was stored in a dataset based on the principles of the Cross-Linguistic Data Formats (Forkel et al. 2018). The 100 most frequent body-object colexifications were added as lexical features to a parameter table so that they are reusable for other researchers. The most frequent body-object colexifications were SKIN-LEATHER (160 language varieties), SKIN-BARK (90 language varieties), and TESTICLE-EGG (31 language varieties). While the study by Tjuka (2024b) included full colexifications, a comparison with previous descriptions of body-object colexifications in the literature and the Database of Semantic Shifts (Zalizniak et al. 2016-2020) revealed that there are potential candidates for affix colexifications in different languages, such as *ʔ ol šil* 'tree skin' for the concept BARK in Yuki.

To identify languages with affix colexifications, we used the network presented in List (2023). The network is based on an updated version of the Intercontinental Dictionary Series (IDS, Borin et al. 2013; Key and Comrie 2023) with word lists covering 1,310 concepts. The language sample in List (2023) included 329 language varieties from 60 language families. With 82 language varieties, the Nakh-Dagestanian language family was the most represented. The Indo-European language family was in second place, with 54 language varieties. The Austroasiatic language family was the third most frequent language family with 37 language varieties. Geographically, there was a bias towards Eurasia with 216 language varieties from this macro area. South America was the second largest macro area, represented by 76 language varieties. The language sample is an opportunistic sample since the number of concepts for each language variety is limited and our analyses are exploratory.

Based on the IDS word lists, List (2023) created a weighted directed network that includes directional tendencies of colexifications of word forms for concepts. The network contains connections between two concepts if an affix colexification of concept A in concept B is detected. For example, *ʔ ol šil* includes the word form *šil* for SKIN, so the algorithm establishes a connection in the direction from SKIN to BARK. If there is a word for SKIN that contains the word form for BARK, then an additional connection in the direction from BARK to SKIN is established. The network represented in List (2023) is available as a graph in GML format and as a concept list representing the network structure in tabular form, using JSON to represent network relations in the cells of the table, which is included in Concepticon (Version 3.2, List et al. 2024).

We calculated the frequency of affix colexifications from body to object concepts and vice versa across languages from List's network using a Python script.[1] The representation of the network as a concept list makes it straightforward to identify languages with an affix colexification between two concepts and the script outputs a list of language frequency counts. To visualize subgraphs of specific body-object colexifications, we used Cytoscape (Shannon et al. 2003), where the nodes represent a concept and the edges (i.e., lines) represent the colexification between two concepts (for a tutorial on how to create colexification networks, see Tjuka 2024a).

3 Results

Of the 100 body-object colexifications in the seed list from Tjuka (2024b), data for 39 body-object colexifications were found in the affix colexification network in List (2023). The most frequent body-object colexifications largely overlapped in both datasets, but less frequent colexifications such as TENDON-ROOT, SKIN-BOOK or TOOTH-LEAF were not present in List (2023).

3.1 Directional tendencies

The frequency across languages with which a body-object colexification occurred either in the direction that the word form for a body concept is used to express the object concept or vice versa, was the basis for the analysis of directionality. Table 1 shows the directional tendencies between body and object concepts including the number of languages with an affix colexification for each direction and the total number of occurrences. The dominant direction was determined by evaluating the counts in each direction, and if there was a direction with at least two instances more in one direction, that direction was determined to be the dominant one. Twenty-one out of 39 colexifications show a directional tendency from body to object (blue arrows), while 16 colexifications show a directional tendency from object to body (orange arrows). In two cases – INTESTINES-SAUSAGE and LIP-SHORE – no directional tendency exists (yellow arrows).

1 The data and code underlying this study are curated on GitHub: https://github.com/calc-project/partial-body-object/tree/v1.0.

Table 1: Directional tendencies in body-object colexifications. The direction from body to object is given in blue, the direction from object to body in orange, and no directional tendencies in yellow.

Body	Direction	Object	→	←	Total
EAR		EARRING	66	2	68
SKIN		BARK	42	6	48
NECK		COLLAR	44	0	44
TONGUE		FLAME	29	0	29
WAIST		BELT	24	5	29
INTESTINES		SAUSAGE	13	14	27
TESTICLES		EGG	2	24	26
FOOT		SHOE	24	0	24
SKIN		LEATHER	18	6	24
SKULL		TOP	0	14	14
LIP		EDGE	3	9	12
SHOULDER BLADE		SPADE	0	12	12
FOOT		WHEEL	11	0	11
TESTICLES		FRUIT	0	10	10
TESTICLES		SEED	0	10	10
HEAD		TOP	6	3	9
BACK		ROOF	8	0	8
SHOULDER BLADE		OAR	0	8	8
SHOULDER BLADE		PADDLE	0	8	8
KIDNEY		SEED	0	7	7
MOUTH		DOOR	5	2	7
NOSE		CAPE	7	0	7
BODY		TREE TRUNK	6	0	6
EYE		SEED	4	2	6
BLOOD VESSEL		ROOT	0	5	5
HEAD		ROOF	5	0	5
THROAT		COLLAR	5	0	5
BUTTOCKS		BOTTOM	0	4	4
HAIR (HEAD)		LEAF	0	4	4
LIP		SHORE	2	2	4
SHOULDER BLADE		SHOVEL	0	4	4
SKIN		SHELL	4	0	4
EYE		FRUIT	0	3	3
FINGERNAIL		NAIL (TOOL)	0	3	3
HAIR (BODY)		LEAF	0	3	3
MOUTH		HOLE	3	0	3
EYE		FIRE	2	0	2
MOUTH		EDGE	2	0	2
TESTICLES		BALL	0	2	2

The most frequent body-object colexification is EAR-EARRING in which the word form for EAR is used to express the concept EARRING in 66 language varieties. Examples include *kula-pepeiao* lit. 'gold-ear' in Hawaiian or *sau faliŋa* lit. 'king ear' meaning EARRING in Rotuman, an Austronesian language. The second most frequent body-object colexification is SKIN-BARK where the word form for SKIN is used to express BARK in 42 language varieties, for example, in Kalamang, a West Bomberai language, *ror kulun* lit. 'tree/wood skin'. The third most frequent body-object colexification with a directional tendency from body to object is NECK-COLLAR, as in *sɨpluw toř* lit. 'neck cloth' in Mansi, a Uralic language, or *ynĩ te*ʔ lit. 'neck clothing' in Chatino, a Zacatepec variety of the Otomanguean language family.

While many different body concepts are used to express object concepts, there are particular body concepts for which the source are object names. The concept TESTICLES has predominantly affix colexifications consisting of names for objects such as EGG, FRUIT, SEED, and BALL. For example, *zaad.bal* lit. 'seed.ball' in Dutch. This directional tendency is due to taboo conventions with respect to reproductive body parts in many cultures. Another body concept that is expressed with words for objects in many languages is SHOULDER BLADE. However, these body-object colexifications are an areal phenomenon that occurs mainly in the Nakh-Daghestanian languages (see also Tjuka 2024b).

Within the 15 most frequent body-object colexifications, which occur in at least ten language varieties, nine body-object colexifications show a directional tendency from body to object concepts. In comparison, only five frequent body-object colexifications show a directional tendency from object to body concepts. It appears that the more frequently a body-object colexification is, the more likely a transfer from body to object occurs. However, this generalization should be treated with caution because in many cases the scarcity of data can lead to infrequent patterns. Object concepts such as SHOVEL or ROOT and body concepts such as BLOOD VESSEL or THROAT are not commonly represented in the word lists. The interpretation of the frequencies must therefore be understood as a tendency rather than a universal.

3.2 Network representations

The discussion of frequencies suggests a slight tendency for body concepts to be the source of object names. However, the picture becomes more complex when examining the network of affix colexifications. Here, we present three subgraphs of the affix network based on List (2023) to illustrate the different ways in which words for body and object concepts are used to express other concepts. In the graphs, body concepts are represented by blue ellipses and object concepts by orange rectangles.

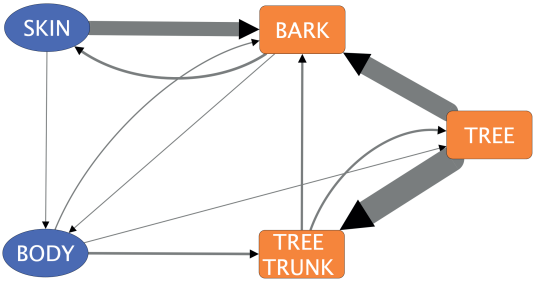

Figure 1: Subgraph with the body concepts SKIN and BODY. Body concepts = blue ellipsis; object concepts = orange rectangle; thickness of edge = frequency across languages; arrow = directionality.

The thickness of the edges indicates the frequency across languages and the arrows indicate the target concepts.

Figure 1 shows the subgraph with the body concepts SKIN and BODY and the object concepts BARK, TREE, and TREE TRUNK. The network reveals that there are many languages in which the word forms for the concepts SKIN and TREE are used for the concept BARK. The word forms for TREE are also frequently used to express TREE TRUNK and for example, in Ende, a Pahoturi language, *llo pätt* lit. 'tree body'. There are other cases in which the word forms for BODY are used for BARK. In these cases, it is likely that SKIN and BODY are colexified and then the word form is used to express BARK. Interestingly, the network shows that there are a few affix colexifications in which the word forms for SKIN express BODY, but not the other way around.

Figure 2 illustrates the different directional tendencies between the body concept TESTICLES and the object concepts SEED, EGG, and FRUIT. Word forms for object concepts are predominantly the basis for expressing the concept TESTICLES, while the word forms for TESTICLES occur less frequently in the expression of the object concepts. In addition, the word forms for the object concepts are also used to express the other object concepts, such as in Hawaiian *hua-ʔai* lit. 'egg-food' for FRUIT. There are also many languages that have full colexifications between the three object concepts.[2] The semantic motivations that lead to particular affix colexifications can therefore not straightforwardly be distinguished.

Figure 3 shows the subgraph with the colexifications between the body concepts MOUTH and LIP and the object concepts EDGE and SHORE. It becomes

2 Compare with: https://clics.clld.org/graphs/subgraph_744.

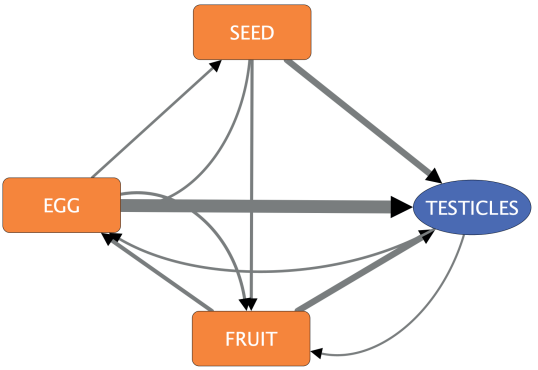

Figure 2: Subgraph with the body concepts TESTICLES. Body concepts = blue ellipsis; object concepts = orange rectangle; thickness of edge = frequency across languages; arrow = directionality.

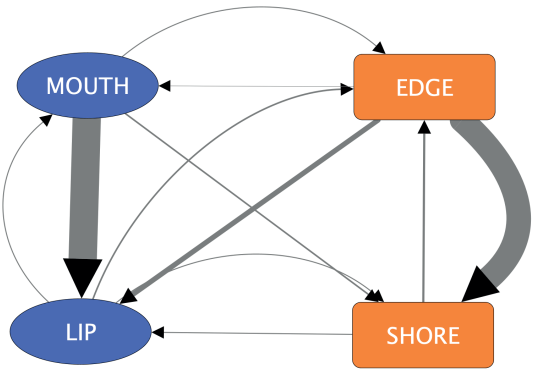

Figure 3: Subgraph with the body concepts MOUTH and LIP. Body concepts = blue ellipsis; object concepts = orange rectangle; thickness of edge = frequency across languages; arrow = directionality.

apparent that the word forms for MOUTH are frequently used to express LIP and that the word forms for EDGE are frequently used to express SHORE. An example of the former is *wal bii* lit. 'leaf mouth' in Polci, an Afro-Asiatic language, and an example of the latter is *wiɲi ʔdanum* lit. 'water.edge' in Wapishana, an Arawakan language. In some cases, the word forms for EDGE are used to express LIP, as in *vom dor̃* lit. 'mouth edge' in Komi, a Uralic language. Although general tendencies arise, there are multiple cases in which word forms are sources for either body or object concepts.

4 Discussion

The result of the study indicates that there is a slight tendency across languages for words expressing body concepts to name object concepts. In many cases, the domain of the human body serves as the source for the target domain of everyday objects. This result is in line with previous proposals that languages systematically use body part terms to refer to objects (Levinson 1994). However, there are almost as many body-object colexifications for which the direction was found to be reversed. In particular, these included the concepts TESTICLES and SHOULDER BLADE, which were named after object concepts more frequently. The former pattern is due to taboo conventions in many cultures, while the latter is an areal pattern in the Nakh-Dagestanian languages. The network representations revealed further variations, including frequent colexifications within the body domain, such as the colexification MOUTH-LIP and within the object domain, as in EDGE-SHORE.

The analysis of affix colexifications was based on existing datasets (List 2023; Tjuka 2024b). While the study by Tjuka (2024b) focused on full colexifications between body and object concepts, the study by List (2023) presented methods for inferring partial colexifications. The present study combined the two approaches and illustrated the synergies that arise through combining data and methods. The investigation of partial colexifications between body and object concepts provides further insights into the semantic motivations that underlie body-object colexifications and completes the patterns found in Tjuka (2024b). As a next step, the partial colexifications between body and object concepts can be used to investigate areal phenomena in which a body-object colexification is expressed by a particular combination of word forms in languages in the same geographical area. Another possibility would be to extend the study of List (2023) with additional datasets and investigate universal tendencies.

One limitation of our study is the scarcity of data. On the one hand, body concepts are in general more commonly featured in word lists because they are part of the basic vocabulary, which is assumed to be universal and stable over time and is therefore used in language comparison studies (Tadmor 2009, 65). Object concepts, on the other hand, are less frequently represented in word lists, so that certain body-object colexifications appear either less frequent than they are or appear as areal phenomena because the concepts are documented for a particular language group, even though the semantic association between the concepts is more widespread. One way to circumvent the scarcity of data is to focus on specific sets of concepts and then conduct targeted data collection across languages. An example of this approach is the study by Norcliffe and Majid (2024), who collected perception verbs in a balanced sample of 100 languages to analyze lexicalization patterns. While the selection of perception verbs was top-down, our study follows

a bottom-up approach. With the list of full body-object colexifications created in Tjuka (2024b) and the partial colexifications found in the present study, targeted data collection can now be conducted. This approach is particularly useful in semantic domains for which only limited research exists and which have not yet been systematically investigated.

5 Outlook

The next step for the study of body-object colexifications is targeted data collection. Not only do we need to collect word forms for the body and object concepts in a balanced sample of languages in order to make more robust predictions about universal tendencies, but we also need to collect data to examine the causes of directional tendencies. There are two possibilities to extend the present study and investigate the causes of colexification.

First, the data in the Database of Semantic Shifts (DatSemShifts, Zalizniak et al. 2024) could be used to analyze patterns in semantic changes between body and object concepts in different languages. This would reveal regularities in semantic change. The data have already been processed and used in a study by Bocklage et al. (2024). However, the glosses in the database are fine-grained and not standardized. For example, the colexification SKIN-BARK cannot be automatically identified because DatSemShifts uses the gloss 'skin (of person)', which is mapped to the narrower concept SKIN (OF HUMAN) in Concepticon (List et al. 2024). In addition, the gloss 'shovel, spade' contains two separate concepts, so that the colexification SHOULDER BLADE-SHOVEL and SHOULDER BLADE-SPADE cannot be detected automatically. The data must be prepared in such a way that a comparison with the body-object colexifications can be performed automatically.

Second, the data in the Cross-Linguistic Database of Norms, Ratings, and Relations for Words and Concepts (NoRaRe, Tjuka et al. 2022, 2023) could be used to investigate correlations between psycholinguistic measures to understand the causes behind body-object colexifications. For example, Xu et al. (2017) showed that psycholinguistic measures predict semantic changes across semantic domains in English. In addition, Winter and Srinivasan (2022) demonstrated that word frequency is a predictor of semantic change between two concrete domains in English. NoRaRe provides data across diverse languages for ratings on concreteness, imageability, arousal, and valence on the one hand and norms on word frequency on the other hand. Here too, the data are still sparse for the concepts examined in this study. Therefore, targeted data collection on these psycholinguistic measures in diverse languages would be desirable. With this type of data, causal predictions

could be systematically tested. This would be an advantage over previous studies, which often rely on researchers' intuitions to interpret a particular relation between two concepts rather than relying on speakers' judgments.

Data availability statement

The data and code that support the findings of this study are openly available on GitHub at https://github.com/calc-project/partial-body-object/tree/v1.0. The data and code are also archived on Zenodo (https://doi.org/10.5281/zenodo.13622990).

Acknowledgments

JML was supported by the ERC Consolidator Grant ProduSemy (PI Johann-Mattis List, Grant No. 101044282, see https://doi.org/10.3030/101044282). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency (nor any other funding agencies involved). Neither the European Union nor the granting authority can be held responsible for them.

References

Baranyiné Kóczy, Judit and Katalin Sipőcz. 2023. *Embodiment in Cross-Linguistic Studies: The 'Heart'*. Leiden: Brill.

Baş, Melike and Iwona Kraska-Szlenk (eds.). 2022. *Embodiment in Cross-Linguistic Studies: The 'Eye'*. Leiden: Brill.

Bocklage, Katja, Anna Di Natale, Annika Tjuka and Johann-Mattis List. 2024. Directional Tendencies in Semantic Change. doi: 10.17613/0y0r-f341.

Borin, Lars, Bernard Comrie and Anju Saxena. 2013. The Intercontinental Dictionary Series – A Rich and Principled Database for Language Comparison. In Lars Borin and Anju Saxena (eds.), *Approaches to Measuring Linguistic Differences*, 285–302. Berlin: De Gruyter Mouton.

Brenzinger, Matthias and Iwona Kraska-Szlenk (eds.). 2014. *The Body in Language – Comparative Studies of Linguistic Embodiment*. Leiden: Brill.

Brochhagen, Thomas and Gemma Boleda. 2022. When Do Languages Use the Same Word for Different Meanings? The Goldilocks Principle in Colexification. *Cognition* 226. 1–8. doi: 10.1016/j.cognition.2022.105179.

Brown, Cecil H. and Stanley R. Witkowski. 1981. Figurative Language in a Universalist Perspective. *American Ethnologist* 8(3). 596–615. doi: 10.1525/ae.1981.8.3.02a00110.

Croft, William. 2016. Typology and the Future of Cognitive Linguistics. *Cognitive Linguistics* 27(4). 587–602. doi: 10.1515/cog-2016-0056.

Dingemanse, Mark. 2009. The Selective Advantage of Body-Part Terms. *Journal of Pragmatics (Special Issue on Communicating Place, Space and Mobility)* 41(10). 2130–2136. doi: 10.1016/j.pragma.2008.11.008.

Forkel, Robert, Johann-Mattis List, Simon J. Greenhill, Christoph Rzymski, Sebastian Bank, Michael Cysouw, Harald Hammarström, Martin Haspelmath, Gereon A. Kaiping and Russell D. Gray. 2018. Cross-Linguistic Data Formats, Advancing Data Sharing and Re-Use in Comparative Linguistics. *Scientific Data* 5(1). 1–10. doi: 10.1038/sdata.2018.205.

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*, 163–215. Amsterdam: John Benjamins. doi: 10.1075/slcs.106.09fra.

Ibarretxe-Antuñano, Iraide. 2012. The Importance of Unveiling Conceptual Metaphors in a Minority Language. In Anna Idström and Elisabeth Piirainen (eds.), *Endangered Metaphors*, 253–272. Amsterdam: John Benjamins. doi: 10.1075/clscc.2.12iba.

Jackson, Joshua Conrad, Joseph Watts, Teague R. Henry, Johann-Mattis List, Robert Forkel, Peter J. Mucha, Simon J. Greenhill, Russell D. Gray and Kristen A. Lindquist. 2019. Emotion Semantics Show Both Cultural Variation and Universal Structure. *Science* 366. 1517–1522. doi: 10.1126/science.aaw8160.

Key, Mary Ritchie and Bernard Comrie. 2023. *The Intercontinental Dictionary Series (Version 4.3)*. Leipzig: Max Planck Institute for Evolutionary Anthropology. doi: 10.5281/zenodo.7701635. http://ids.clld.org.

Koch, Peter. 2008. Cognitive Onomasiology and Lexical Change: Around the Eye. In Martine Vanhove (ed.), *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations*, 107–137. Amsterdam: John Benjamins. doi: 10.1075/slcs.106.07koc.

Koch, Peter and Daniela Marzo. 2007. A Two-Dimensional Approach to the Study of Motivation in Lexical Typology and Its First Application to French High-Frequency Vocabulary. *Studies in Language* 31(2). 259–291. doi: 10.1075/sl.31.2.02koc.

Kraska-Szlenk, Iwona. 2014. Semantic Extensions of Body Part Terms: Common Patterns and Their Interpretation. *Language Sciences* 44. 15–39. doi: 10.1016/j.langsci.2014.02.002.

Kraska-Szlenk, Iwona (ed.). 2020. *Body Part Terms in Conceptualization and Language Usage*. Amsterdam: John Benjamins. doi: 10.1075/clscc.12.

Lakoff, George and Mark Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

Lakoff, George and Mark Johnson. 2003-[1980]. *Metaphors We Live By*. Chicago: University of Chicago Press revised edition edn.

Levinson, Stephen C. 1994. Vision, Shape, and Linguistic Description: Tzeltal Body-Part Terminology and Object Description. *Linguistics* 32(4-5). 791–855. doi: 10.1515/ling.1994.32.4-5.791.

List, Johann-Mattis. 2023. Inference of Partial Colexifications from Multilingual Wordlists. *Frontiers in Psychology* 14. 1–10. doi: 10.3389/fpsyg.2023.1156540.

List, Johann-Mattis, Robert Forkel, Simon J. Greenhill, Christoph Rzymski, Johannes Englisch and Russell D. Gray. 2022. Lexibank, a Public Repository of Standardized Wordlists with Computed Phonological and Lexical Features. *Scientific Data* 9(1). 316. doi: 10.1038/s41597-022-01432-0.

List, Johann-Mattis, Simon J. Greenhill, Cormac Anderson, Thomas Mayer, Tiago Tresoldi and Robert Forkel. 2018. CLICS^2: An Improved Database of Cross-Linguistic Colexifications Assembling Lexical Data with the Help of Cross-Linguistic Data Formats. *Linguistic Typology* 22(2). 277–306. doi: 10.1515/lingty-2018-0010.

List, Johann Mattis, Annika Tjuka, Mathilda van Zantwijk, Frederic Blum, Carlos Barrientos Ugarte, Christoph Rzymski, Simon Greenhill and Robert Forkel (eds.). 2024. *CLLD Concepticon 3.2.0*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://concepticon.clld.org/.

Mayer, Thomas, Johann-Mattis List, Anselm Terhalle and Matthias Urban. 2014. An Interactive Visualization of Crosslinguistic Colexification Patterns. In Annette Hautli-Janisz, Verena Lyding and Christian Rohrdantz (eds.), *Proceedings of the LREC Workshop 'VisLR: Visualization as Added Value in the Development, Use and Evaluation of Language Resources'*, 1–8. Reykjavik, Iceland: European Language Resources Association.

Norcliffe, Elisabeth and Asifa Majid. 2024. Verbs of Perception: A Quantitative Typological Study. *Language* 100(1). 81–123.

Pattillo, Kelsie E. and Małgorzata Waśniewska (eds.). 2023. *Embodiment in Cross-Linguistic Studies: The 'Face'*. Leiden: Brill.

Ponsonnet, Maïa. 2014. Figurative and Non-Figurative Use of Body-Part Words in Descriptions of Emotions in Dalabon (Northern Australia). *International Journal of Language and Culture* 1(1). 98–130. doi: 10.1075/ijolc.1.1.06pon.

Ponsonnet, Maïa and Kitty-Jean Laginha. 2020. The Role of the Body in Descriptions of Emotions: A Typology of the Australian Continent. *Pragmatics & Cognition* 27(1). 20–82. doi: 10.1075/pc.00011.pon.

Rzymski, Christoph, Tiago Tresoldi, Simon J. Greenhill, Mei-Shin Wu, Nathanael Schweikhard, Maria Koptjevskaja-Tamm, Volker Gast, Timotheus A. Bodt, Abbie Hantgan, Gereon A. Kaiping, Sophie Chang, Yunfan Lai, Natalia Morozova, Heini Arjava, Nataliia Hübler, Ezequiel Koile, Steve Pepper, Mariann Proos, Briana Van Epps, Ingrid Blanco, Carolin Hundt, Sergei Monakhov, Kristina Pianykh, Sallona Ramesh, Russell D. Gray, Robert Forkel and Johann-Mattis List. 2020. The Database of Cross-Linguistic Colexifications, Reproducible Analysis of Cross-Linguistic Polysemies. *Scientific Data* 7(1). 1–12. doi: 10.1038/s41597-019-0341-x.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski and Trey Ideker. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* 13(11). 2498–2504. doi: 10.1101/gr.1239303.

Steinberg, Reinhild. 2014. *Lexikalischer Wandel und Polygenese: Der Konzeptbereich KOPF*. Tübingen: Stauffenburg-Verlag Narr. (Doctoral Dissertation).

Tadmor, Uri. 2009. Loanwords in the World's Languages: Findings and Results. In Martin Haspelmath and Uri Tadmor (eds.), *Loanwords in the World's Languages*, 55–75. Berlin: De Gruyter Mouton. doi: 10.1515/9783110218442.55.

Tjuka, Annika. 2024a. How to Visualize Colexification Networks in Cytoscape (How to Do X in Linguistics 14). *Computer-Assisted Language Comparison in Practice* 7(1). 7–16. doi: 10.15475/calcip.2024.1.2. https://calc.hypotheses.org/6697.

Tjuka, Annika. 2024b. Objects as Human Bodies: Cross-Linguistic Colexifications Between Words for Body Parts and Objects. *Linguistic Typology* 1–40. doi: 10.1515/lingty-2023-0032.

Tjuka, Annika, Robert Forkel and Johann-Mattis List. 2022. Linking Norms, Ratings, and Relations of Words and Concepts Across Multiple Language Varieties. *Behavior Research Methods* 54. 864–884. doi: 10.3758/s13428-021-01650-1. https://norare.clld.org/.

Tjuka, Annika, Robert Forkel and Johann-Mattis List. 2023. Curating and Extending Data for Language Comparison in Concepticon and NoRaRe. *Open Research Europe* 2(141). 1–13. doi: 10.12688/openreseurope.15380.3.

Traugott, Elizabeth Closs and Richard B. Dasher. 2001. *Regularity in Semantic Change*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486500.

Urban, Matthias. 2011. Asymmetries in Overt Marking and Directionality in Semantic Change. *Journal of Historical Linguistics* 1(1). 3–47. doi: 10.1075/jhl.1.1.02urb.

Wilkins, David P. 1996. Natural Tendencies of Semantic Change and the Search for Cognates. In Mark Durie and Malcolm Ross (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, 264–304. Oxford: Oxford University Press.

Winter, Bodo and Mahesh Srinivasan. 2022. Why Is Semantic Change Asymmetric? The Role of Concreteness and Word Frequency and Metaphor and Metonymy. *Metaphor and Symbol* 37(1). 39–54. doi: 10.1080/10926488.2021.1945419.

Xu, Yang, Khang Duong, Barbara C. Malt, Serena Jiang and Mahesh Srinivasan. 2020. Conceptual Relations Predict Colexification Across Languages. *Cognition* 201. 1–9. doi: 10.1016/j.cognition.2020.104280.

Xu, Yang, Barbara C. Malt and Mahesh Srinivasan. 2017. Evolution of Word Meanings Through Metaphorical Mapping: Systematicity Over the Past Millennium. *Cognitive Psychology* 96. 41–53. doi: 10.1016/j.cogpsych.2017.05.005.

Yu, Ning. 2002. Body and Emotion: Body Parts in Chinese Expression of Emotion. *Pragmatics & Cognition* 10(1-2). 341–367. doi: 10.1075/pc.10.12.14yu.

Zalizniak, Anna A., Anna Smirnitskaya, Maksim Russo, Ilya Gruntov, Timur Maisak, Dmitry Ganenkov, Maria Bulakh, Maria Orlova, Marina Bobrik-Fremke, Oksana Dereza, Tatiana Mikhailova, Maria Bibaeva and Mikhail Voronov (eds.). 2016–2020. *Database of Semantic Shifts*. Moscow: Institute of Linguistics, Russian Academy of Sciences. http://datsemshift.ru.

Zalizniak, Anna A., Anna Smirnitskaya, Maksim Russo, Ilya Gruntov, Timur Maisak, Dmitry Ganenkov, Maria Bulakh, Maria Orlova, Marina Bobrik-Fremke, Oksana Dereza, Tatiana Mikhailova, Maria Bibaeva and Mikhail Voronov (eds.). 2024. *Database of Semantic Shifts*. Moscow: Institute of Linguistics, Russian Academy of Sciences. http://datsemshift.ru.