Annika Tjuka*

Objects as human bodies: cross-linguistic colexifications between words for body parts and objects

https://doi.org/10.1515/lingty-2023-0032
Received April 9, 2023; accepted February 17, 2024; published online April 11, 2024

Abstract: Many languages have words that denote a human body part and an object, for example, *hand*, which refers to a part of a person and a watch. As of yet, there is no systematic study on the distribution of these shared names, i.e., colexifications, between two concrete semantic domains in a variety of languages. Here, I present a study that investigates colexifications between body and object concepts, i.e., body-object colexifications. By using a newly established workflow, colexifications are automatically extracted based on a seed list containing 134 body concepts and 650 object concepts. The analysis focuses on the frequency, distribution, cognitive relations, and coincidental cases of 78 body-object colexifications occurring across 396 language varieties. The results show that some body-object colexifications are widespread, but most occur in a small number of language varieties. By creating a network structure to examine individual relations and additionally comparing ratings of visual and haptic perception across concepts, the study indicates that the similarity of visual perception plays a central role in the emergence of body-object colexifications. The findings provide a first general overview of the phenomenon and offer ample opportunities for future research.

Keywords: lexical typology; body parts; objects; colexification; cognition

1 Introduction

Examples of using the same word for a body part and an object can be found across many languages. For instance, Western Apache (west2615) extends the meanings of body part terms systematically to cars: *wos* ‘shoulder, front fender(s)’, *inda’* ‘eye(s), headlights’, *pıt* ‘stomach, gas tank’, etc. (Basso 1967: 472). Similarly, speakers of diverse languages use body part terms to describe objects in the landscape, as in

***Corresponding author: Annika Tjuka** [ˈɑnːikɑ tjʊˈka], Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany, E-mail: annika_tjuka@eva.mpg.de

Marquesan (marq1246) where *ivi* ‘bone’ also means ‘mountain range peak, hill’ (Cablitz 2008: 205), or *punth* in Kuuk Thaayorre (thay1249), which can refer to a human arm and a tree branch (Gaby 2006: 218). Additional examples appear in studies that examine universal and areal patterns of lexico-semantic associations from a synchronic or diachronic perspective (e.g., Brown and Witkowski 1981, 1983; Gast and Koptjevskaja-Tamm 2019; Urban 2012; Wilkins 1996). One of the most exhaustive studies analysing the lexico-semantic associations of 19 body part terms found in dictionaries of 149 languages is the dissertation by Urban (2012). Furthermore, the dissertation by Steinberg (2014) offers detailed descriptions of semantic changes and typological patterns of body part terms in the domain of HEAD. However, none of the previous studies focused exclusively on the shared names between body parts and objects.

The present study examines the relation between the human body and objects by analysing “strict colexifications” (François 2008: 10).[1] A strict colexification occurs when the same lexical form is used to denote two senses based on a synchronic analysis, whereas “loose colexifications” include links between senses from a diachronic perspective and links based on derivation or composition (François 2008: 10). The expression *river mouth* is thus a loose colexification and *punth* ‘arm, branch’ in Kuuk Thaayorre (Gaby 2006: 218) is a strict colexification. If a language has two separate lexical forms for ARM and BRANCH, the two senses are “dislexified” (François 2021: 5), as in German (stan1295) and English (stan1293): *Arm* ‘arm’ and *Ast* ‘branch’. I use the term “body-object colexifications” for instances in which a body part and an object share the same name.[2] The objects include inanimate things in our surroundings that can be perceived visually and/or haptically: food, clothing items, landscape features, and parts of objects like edges or sides. Although “body” appears first, the term body-object colexification does not imply a directionality of body concepts as source and object concepts as a target.

The focus on empirical data in the analysis of colexifications has led to methodological advances such as the creation of the Cross-Linguistic Database of Colexifications (CLICS[3] Rzymski et al. 2020). CLICS in turn was the initial step for the collection of standardised word lists in the Cross-Linguistic Data Formats (CLDF, Forkel et al. 2018) and was later extended into a large collection of lexical data:

1 The term “full colexification” in contrast to “partial colexification” has been introduced (List 2023; List et al. 2022), but I adhere to the terminology by François (2008) in this article.

2 The reason for using the term “colexification” rather than polysemy is twofold. First, the term “colexification” unites instances of polysemy, homonomy, and vagueness which can be difficult to differentiate even in an individual language (cf. Behrens 2002; Geeraerts 1993; Lehrer 1990; Tuggy 1993). Second, the use of empirical cross-linguistic observations as the basis for distinguishing meanings as proposed by François (2008: 19) has the advantage of identifying lexical patterns arising from the data rather than pre-selecting them in advance.

Lexibank (List et al. 2022). In addition to providing lexical data in a standardised format, Lexibank offers more flexibility in extracting colexifications from the data. Based on a newly established workflow using a set of 36 data sets from Lexibank, the study analyses 78 strict body-object colexifications occurring across 396 language varieties.

While the study gives an overview of body-object colexifications across diverse languages, qualitative studies can provide further insights through consultations and contextual examples (e.g., Schapper 2022; Souag 2022). I therefore include additional resources such as the Database of Semantic Shifts (Zalizniak et al. 2016) and examples from studies on the meanings of body part terms (e.g., Brown and Witkowski 1981, 1983; Hilpert 2007; Steinberg 2014; Urban 2012; Wilkins 1996) to discuss the results. The goal is to examine the frequency, distribution, and cognitive relations of body-object colexifications. The main research question is: How widespread are body-object colexifications across languages and what are the causes for the emerging patterns? Since there are no studies of similar scope, the study is to some extent exploratory.

The study adopts a data-driven rather than a theory-driven approach. However, the study contributes to the theoretical discussion on similarity and salience. First, while body-object colexifications are often categorised as metaphors, their systematic use in some languages challenges the assumption that they are metaphorical extensions (Levinson 1994). Terminological differentiation between metaphor and metonymy remains controversial (cf. Goossens 1990; Riemer 2002) and the linguistic variation hinders unambiguous categorisation. Rather than organising body-object colexifications under these theoretical constructs, I argue that they fall on a scale ranging between figurative and literal similarity (cf. Geeraerts 2015: 425). Second, the argument that salience drives colexifications of certain body parts (Andersen 1978; Brown 1976; Kraska-Szlenk 2014) faces two caveats: (a) salient body parts are defined *a priori* so that evidence is collected for the choices made; (b) salience is measured by the researcher's intuition. The study addresses these limitations by analysing a large number of body part concepts and comparing English speakers' ratings of vision and touch for body and object concepts.

2 Body part terms and their meanings

The semantic domain of the human body, alongside colour and kinship, is extensively studied in lexical typology (Koptjevskaja-Tamm et al. 2007: 161). Cross-linguistic studies have examined how languages name the parts of the body (Andersen 1978; Brown 1976) and studies using non-linguistic stimuli have added further insights into naming patterns (Huisman et al. 2021; Majid 2010; Majid et al.

2006). Only two studies have systematically explored body-object colexifications in two Mesoamerican languages and English (Levinson 1994; Tilbe 2017). However, many examples are mentioned within the broader context of lexico-semantic associations. Here, I discuss studies on colexifications and the four interacting factors that shape their patterns: language contact, historical processes, cognitive principles, and coincidence (Koptjevskaja-Tamm and Liljegren 2017: 224–229).

2.1 Areal distribution of body-object colexifications

Lexical typology is closely connected to areal semantics, which aims to explain the diffusion of lexical patterns due to language contact (Koptjevskaja-Tamm and Liljegren 2017: 204). Studies in areal semantics list body-object colexifications as a lexical pattern specific to a linguistic area (e.g., Campbell et al. 1986; Gast and Koptjevskaja-Tamm 2019; Souag 2022; Urban 2010). Here, the distinction between strict and loose colexifications is particularly relevant. Since the same lexico-semantic association between a body part and object can be expressed by a strict colexification as in Lao (laoo1244) *paak5* 'mouth (of a bottle, a river)' (Enfield 2006: 193) or as a loose colexification such as English *river mouth*, a qualitative analysis of the languages in a given geographic area is required to establish a recurrent lexical pattern. For example, Urban (2010) identifies expressions with the meaning *eye of day* for the concept SUN as an areal lexical pattern occurring in languages belonging to the Austroasiatic, Tai-Kadai, and Austronesian language families. Arguing against the proposed areality of *eye of day*, Blust (2011) presents reconstructions of Austronesian languages and expressions for *eye of sun* to show that the pattern is more widespread. Thus, different realisations can stand for the same lexico-semantic association between a body part and an object. It is however also possible that a lexico-semantic association is widespread in diverse languages, while a particular expression, i.e., a loose colexification, is specific to a language group in a geographical area.

Details about these different expressions of lexico-semantic associations are often passed over in studies using lexical databases (Schapper 2022: 3). A lexical database such as CLICS³ (Rzymski et al. 2020) cannot account for the variation of a loose body-object colexification because it includes only strict colexifications. Thus, the colexification between EYE and SUN is listed for four language varieties in CLICS³: Isnag (isna1241) spoken in the Philippines, Puruborá (puru1264) spoken in Brazil, Korupun-Sela (koru1245) and Nalca (nalc1240) both spoken in Indonesia.[3] Due to the limited detail provided by the automated analysis of colexifications in word lists, I expect to find few areal patterns in the study.

3 See https://clics.clld.org/edges/1248-1343.

Table 1: Strict body-object colexifications and the area in which they are proposed to occur.

Body-object colexification	Area	Study
FOOT-WHEEL	South America	List et al. (2018)
EAR-LEAF	Eastern Africa, Americas, Australia	Gast and Koptjevskaja-Tamm (2019)
SKIN-BARK	South America, Melanesia	Gast and Koptjevskaja-Tamm (2019)

Nevertheless, an investigation of strict body-object colexifications is useful, as pointed out by Koptjevskaja-Tamm and Liljegren (2017: 224): "A systematic large-scale investigation gives an important evaluation of how frequent or unique a particular property is among the world's languages and provides an indication of the areas in which it might be found." An example of such a large-scale study is Gast and Koptjevskaja-Tamm (2019) who use data from two lexical databases: Automated Similarity Judgment Program (ASJP) Version 17 (Wichmann et al. 2016) and CLICS Version 1 (List et al. 2013). While the clusters arising in the CLICS[1] data did not include any body-object colexifications, two clusters appeared in the ASJP data: EAR-LEAF and SKIN-BARK. With additional data in CLICS[2], List et al. (2018) found an areal pattern of FOOT-WHEEL. In the ASJP data, the body-object colexification EAR-LEAF occurs mainly in Eastern Africa with weak clusters in the Americas and Australia. SKIN-BARK clusters more strongly in South America and arises in a smaller cluster in Melanesia. This finding is in contrast to the claim by Smith-Stark (1994) that the colexification between SKIN and BARK is a lexical pattern found exclusively in Mesoamerican languages. Thus, large-scale studies can be used to test the validity of an areal pattern. The expectation for the present study is consequently that some of the proposed areal patterns summarised in Table 1 turn out to be more widespread due to the larger data coverage.

2.2 Semantic changes involving terms for body parts and objects

A semantic change occurs when the use of a word is altered. For example, the Proto-Germanic term **kuppa-* referred to 'vessel' and developed into German *Kopf* meaning 'head' (List 2014: 35). While it was long assumed that semantic change occurs randomly (Hock and Joseph 1996/2019: 215), works by Traugott and Dasher (2001) and Wilkins (1996) suggest that semantic changes are in some cases more regular than claimed and that cross-linguistic tendencies can be detected. Regular changes in body part terminology involve an intermediate stage in which a word extends its meaning and becomes polysemous. Thus, "[s]ynchronic polysemy becomes crucial in the investigation of semantic changes because it acts as a proof of the plausibility that two meanings are semantically related and that one meaning could give rise to the other" (Wilkins 1996: 269). In other words, the study of synchronic lexical patterns is connected to the study of semantic change. It is therefore essential for a lexical typological study to include evidence about historical processes to understand the emerging patterns in the world's languages (Koptjevskaja-Tamm et al. 2015: 14). Here, I focus on widespread polysemy patterns and semantic changes involving the words that refer to body parts and objects. Note that the semantic changes described in the literature occur from a body part term to an object term or vice versa.

Brown and Witkowski (1981) demonstrate that lexical forms for EGG frequently extend their meaning to TESTICLE. The polysemy egg/testicle is expressed as a strict colexification in most languages (Brown and Witkowski 1981: 604). In addition, the lexical forms for other round small objects such as STONE, SEED, or FRUIT extend their meaning to TESTICLE although these patterns are less common (cf. Brown and Witkowski 1981; Wilkins 1996). The authors argue that the patterns occur due to the salience of the round shape of the body part. Similar claims can be found for the polysemy patterns eye/seed and eye/fruit (Brown and Witkowski 1983: 77). However, Wilkins (1996) states that "we must take care not to use semantic changes as our only evidence of relative salience if we are also going to use this as our explanation of those same changes." (Wilkins 1996: 279). To find a general tendency regarding salience, the selection of concepts thus needs to be broad.

Semantic change and cross-linguistic patterns of polysemy are often explained by cognitive processes. By combining diachronic and typological data with insights from cognitive linguistics, the framework of diachronic cognitive onomasiology provides an integrated view of meaning extensions (cf. Blank and Koch 1999; Koch 2001, 2016; Steinberg 2014). In a study of the concept EYE, Koch (2008: 128) shows that the terms for EYEBALL are frequently based on terms for round objects such as EGG or BALL across languages. Furthermore, Steinberg (2014) provides an extensive analysis of the domain of HEAD. Steinberg (2014) shows that continuous relations are involved in meaning extensions which indicates that metonymy is a fundamental cognitive relation (see also Koch 2011: 300). This finding supports the observation by Wilkins (1996: 272) who found a tendency for metonymic semantic changes in Dravidian languages involving clothing items, as in the change from FOOTLING to FOOT, EARRING to EAR, and PUBIC TASSEL to PENIS. Thus, frequent patterns of polysemous terms referring to body parts and objects, i.e., clothing items, may be based on metonymic extensions.

2.3 Cognitive relations between body parts and objects

The interpretation of a strict colexification between TESTICLE and EGG as a figurative expression based on their round shape suggests a metaphorical semantic relation (Brown and Witkowski 1981: 601). Conversely, the semantic change from EARRING to EAR demonstrates a metonymic relation based on contiguity (Wilkins 1996: 272). The use of *gɨ̀k yȯˀo* lit. 'head house' meaning ROOF in Zapotec (zapo1437) (MacLaury 1989: 135) is based on a part-of relation, indicating a meronymy. These examples illustrate three cognitive processes – similarity, contiguity, and part-of – underlying body-object colexifications, which reflect the semantic relations metaphor, metonymy, and meronymy.

As of yet, no systematic study of cognitive relations leading to body-object colexifications across diverse languages exists. While the framework of embodiment (cf. Brenzinger and Kraska-Szlenk 2014; Kraska-Szlenk 2020) and conceptual metaphor (cf. Lakoff 1987/1990; Lakoff and Johnson 1980/2003) seem relevant to the scope of the present study, I chose a different approach. The reasons are twofold. First, cognitive linguistics typically assumes a directional mapping from concrete to abstract domains, but this study focuses on two concrete domains, a less common subject of study.[4] Additionally, examples where object names extend their meaning to body parts[5] challenge the embodiment hypothesis, which posits that our mental representations are based on the sensorimotor perception of our body: "The same neural and cognitive mechanisms that allow us to perceive and move around also create our conceptual systems and modes of reason." (Lakoff and Johnson 1999: 18). Second, instances like *foot of the mountain* or *leg of the table* are disregarded in studies on conceptual metaphors because they are not systematic and idiosyncratic in language and thought (Lakoff and Johnson 1980/2003: 54). However, I argue that body-object colexifications offer insights into similarity, contiguity, and part-of relations leading to polysemous words.

Similarity is the cognitive relation underlying metaphor (Geeraerts 2010: 283). Ullmann (1963) states: "Since metaphor is based on the perception of similarities, it is only natural that, when an analogy is obvious, it should give rise to the same metaphor in various languages; hence the wide currency of expressions like the

---

4 An exception is the study by Winter and Srinivasan (2022) who analysed the semantic changes found in Urban (2011) showing that word frequency is a better predictor for semantic changes in concrete domains.

5 Examples include semantic changes in which terms for objects develop into terms for HEAD, EYE, or MOUTH (cf. Steinberg 2014). Wilkins (1996: 272) reports family-specific semantic changes, for example, terms for NEST extending their meaning to BELLY in Dravidian, CAVE to BELLY in Tibeto-Burman, BAG to BELLY in Bantu, and BASKET to BELLY in Papuan languages.

'foot of a hill' or the 'leg of a table'." (Ullmann 1963: 188–189). However, defining "obvious similarity" proves challenging. While visual features are commonly assumed to contribute to similarity and salience, experimental studies suggest functional salience also plays a role (Morrison and Tversky 2005; Tversky and Hemenway 1984). Tjuka (2019) explores different dimensions of similarity, i.e., shape, spatial orientation, and function, as the basis for loose body-object colexifications across 13 languages. The results show that a body part term is more often used for an object when the body part and object are similar in more than one dimension. The interpretation of the factors shaping body-object colexifications relates to the distinction between figurative and literal similarity, but this has not been considered in any theoretical framework yet (Geeraerts 2015: 425). Body-object colexifications tend to be based on literal rather than figurative similarity. I therefore include a case study examining the interaction between ratings of vision and touch across body and object concepts to supplement subjective interpretations of cognitive relations (Section 4.4).

Contiguity is the cognitive relation underlying metonymy (Geeraerts 2010: 282). There are only a few examples of body-object colexifications mentioned in the literature that involve metonymy. Wilkins (1996: 272) and Steinberg (2014: 283) list examples of semantic changes for clothing items that arose from body part terms (Section 2.2). Body-object colexifications are also involved in metonymic chains (Hilpert 2007). The analysis of extensions of body part terms in a corpus of 76 languages shows frequent cross-linguistic extensions such as 'arm' extending to 'hand' (31 languages), 'sleeve' (12 languages), and 'shoulder' (five languages, Hilpert 2007: 8). In addition, common serial extensions based on metaphoric and metonymic mappings such as the path from OBJECTS ARE HUMAN BEINGS > PART FOR ORIENTATION > PLACE FOR ACTION expressed in *back* → *back part* → *behind* are found across languages (Hilpert 2007: 15).

Part-of relations are the cognitive relation underlying meronymy (Cruse 1986: 159). While some languages do not use partial hierarchies to structure their body lexicon (e.g., Kuuk Thaayorre, Gaby 2006), other languages have a strong preference for part-of relations (e.g., Tzeltal (tzel1254) and Zapotec, Tilbe 2017). Part-of relations are closely connected to spatial orientation. In an experimental study, English-speaking preschool children were asked to map six body part terms onto a picture of a tree and two body part terms of the face to a picture of a mountain in different spatial orientations. The children were able to preserve the spatial orientation in referring to a part in the picture (Gentner 1977: 5). Although the present study includes words referring to the parts of a tree, house, and mountain, it is not designed to confirm a full paradigm of partial hierarchies for particular objects.

2.4 Coincidence in the emergence of body-object colexifications

The majority of studies concerned with colexifications analyse patterns of polysemy. Similarly, historical linguists focus on naturally occurring semantic changes rather than coincidental cases (Behrens 2002: 323–324). Coincidental cases in which the same lexical form has two unrelated meanings are linguistic phenomena (i.e., homonymy) that evolve by different chains of events. Thus, they are not immediately obvious and there is a tendency to misinterpret certain connections. For instance, Tagalog (taga1270) borrowed *bangko* from Spanish (stan1288) *banco* and the meanings 'bench' and 'financial institution' were preserved (Behrens 2002: 324). Another example is the use of *ear* in *ear of corn* which was falsely interpreted as being based on a similarity in shape by Tjuka (2019). The similarity is a coincidence since the term evolved from Proto-Germanic **akhuz* 'spike'. English developed a homonymy with *ear*, whereas the difference is still represented in German *Ähre* 'spike' (Behrens 2002: 323).

In lexical databases such as CLICS^3 (Rzymski et al. 2020), restrictions on the networks are implemented to sieve out those coincidental cases and establish patterns of polysemy rather than cases of homonymy (List et al. 2013: 2–3). CLICS^3 includes a threshold so that the network only shows colexifications that occur in at least three language varieties from different language families (Rzymski et al. 2020: 5). Therefore, cross-linguistic colexification networks can only be used to explore general processes across diverse languages rather than specific cases of coincidence. Nevertheless, closely related languages could have adopted or copied a homonymy leading to a pattern across one language group.

3 A cross-linguistic study of body-object colexifications

3.1 Expectations

Based on the literature presented in the previous section, I formulated expectations for the study. They are summarised below:

E1. Even though some areal patterns are to be expected, their emergence in the analysis is less likely due to the coarse-grained approach using strict colexifications.

E2. Since the present study uses a larger data collection than previous studies, some proposed areal patterns may turn out to be more widespread.

E3. Words referring to more salient body parts will tend to colexify more frequently.
E4. Metonymic relations between body parts and objects, specifically clothing items, are likely to occur.
E5. If more than one cognitive relation can be identified as the basis for a body-object colexification, it should occur more frequently in the data.
E6. Only a few coincidental colexifications will appear and if so, they will be restricted to language varieties of the same language group.

Given these expectations, the study investigates factors involved in the emergence of body-object colexifications. While it is beyond the scope of the study to provide a qualitative analysis of loose colexifications, it presents many body-object colexifications that can be explored in future studies. The aim is to conduct the first large-scale study investigating the patterns and causes of strict body-object colexifications.

3.2 Language sample

The study is based on a sample of 931 languages belonging to 137 language families that are spoken across the globe, as shown in Figure 1. The map indicates that it is not a balanced sample and that data from linguistically diverse areas such as Africa, Australia, and Papua New Guinea are missing. The data are based on different sources of which the International Dictionary Series (Key and Comrie 2023) is the largest one. A list of all datasets included in this study can be found in Supplementary Table 1.

The language sample is an opportunistic sample, commonly used in typological studies due to practical motivations (Bakker 2010: 106). The reasons for choosing an opportunistic rather than a probability sample or a variety sample are threefold.

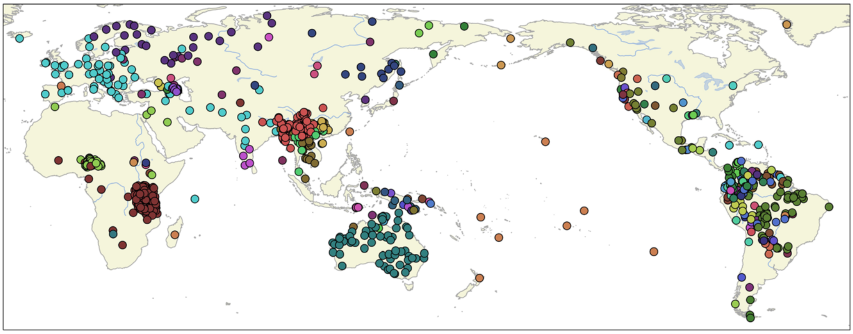

Figure 1: Distribution of all languages in the sample. The 22 largest language families of the 137 are highlighted in distinct colours.

First, the study focuses on a lexical pattern that may not be frequent across all languages. Since it is not clear yet in which language variety more or less body-object colexifications occur, the language sample includes lexical data from all available language varieties to establish frequencies that can be tested with balanced samples in future studies. Second, the number of concepts available in each of the language varieties differs greatly. Some datasets include up to 1,000 concepts; others only about 400. It is therefore not possible to predict which language varieties will have sufficient coverage of lexical forms for body and object concepts. Third, since the study focuses on strict colexifications, the number of attested body-object colexifications is smaller than if loose colexifications were included. So a larger language sample can provide hints on where to look for loose colexifications.

3.3 Materials

The starting point for the analysis of body-object colexifications was selecting a list of "comparative concepts" (Haspelmath 2010) from the semantic domain of the human body and the semantic domain of objects. I used the standardised concepts in the Concepticon (List et al. 2016) which is a reference catalogue for comparative concepts (cf. List et al. 2016; Tjuka et al. 2023). The Concepticon includes word lists such as the 100- and 200-item Swadesh list (Swadesh 1955, 1952) or the Leipzig-Jakarta list (Tadmor 2009). Based on the lexical forms found in these lists, the editors of the Concepticon established standardised concepts consisting of a unique identifier. An example is "1256 HEAD" with the semantic field[6] 'The body' and ontological category[7] 'Person/Thing'. Lexical forms such as *head*, *Kopf*, and *tête* are mapped to 1256 HEAD. The Concepticon currently consists of more than 3,000 concepts and is constantly growing. For the selection of the concepts, I used Concepticon Version 2.5 (List et al. 2021).

The body concepts were selected from the semantic field of 'The body'. In a blog post (Tjuka 2020), I describe an initial list of 171 human and animal body part concepts that were tagged for *category* (human/animal), *gender* (male/female), *part of*, and *instance of*. From this list, I chose 134 human body concepts and excluded 37 animal body concepts.[8] The resulting list consisted of concepts for large human body

6 The semantic fields are based on the categories in the Intercontinental Dictionary Series (Borin et al. 2013) which were taken from the categories in Buck (1949/1988).

7 The ontological categories were adapted from the World Loanword Database (List et al. 2016: 2394). A detailed description of the Concepticon is provided in List et al. (2016) and Tjuka et al. (2023).

8 The distinction between an animal and a human body part is not always unambiguous and colexifications frequently occur. The distinction presented here is based on the mappings in the Concepticon (List et al. 2016), which rely on the information provided by the creators of the underlying word lists.

Table 2: Categories of object concepts, number of Concepticon concepts, and examples.

Category	No. concepts	Examples
Tool	169	945 SPEAR, 1074 FISHHOOK, 2680 MACHETE
Human body part	134	478 LIP, 1303 FINGER, 1291 BACK
Landscape	106	3463 RIVERBANK, 3810 BOULDER, 2423 CHANNEL
Food	95	191 NUT, 593 POTATO, 1569 CHEESE
Household items	95	1254 HANDLE, 1539 BASKET, 2319 LID (COVER, CAP)
Clothing	64	2804 AMULET, 3349 WRISTWATCH, 1680 HELMET
Plant	50	124 THORN, 344 TREE TRUNK, 628 LEAF
Vehicle	28	234 PADDLE, 1597 AIRPLANE, 2622 ENGINE
Spatial relation	19	863 EDGE, 1753 TOP, 682 SIDE
Musical instrument	15	300 RATTLE, 2681 LUTE, 2590 HARMONICA
Animal	9	1065 SPIDER WEB, 3857 EGGSHELL, 3800 IVORY

parts such as 1256 HEAD, 1673 ARM, 1297 LEG, and small body parts such as 1248 EYE, 1380 TOOTH, and 803 ANKLE.

The object concepts were selected from different semantic fields, for example, 'Agriculture and vegetation', 'Clothing and grooming', 'Modern world', and 'The house'. I first filtered the concepts by semantic fields and then I selected the concepts with the ontological category 'Person/Thing' in each semantic field. In this subset, I identified objects and object parts that represent a physical entity. The objects include inanimate things that can be perceived visually and/or haptically such as food or clothing items, objects in the landscape including trees and bodies of water, and parts of objects like edges or sides. To get a better overview of the different object concepts, I grouped them into ten categories. Examples for each category and the number of concepts are given in Table 2. To obtain a broad overview, a large number of object concepts were chosen at the outset.

The resulting seed list of concepts from Concepticon includes 134 body concepts and 650 object concepts (cf. Tjuka 2022). The combination of these concepts gives 87,100 possible body-object colexifications, of which only a fraction is attested.

3.4 Methods

The study is based on a computer-assisted approach that extracts strict colexifications from word lists. The word lists are stored in Lexibank, a large-scale collection of lexical data (List et al. 2022). The data used to create CLICS[3] (Rzymski et al. 2020) is included in Lexibank under the subset *ClicsCore* (List et al. 2022: 5). Of the ClicsCore datasets, I selected 36 datasets with at least 250 items to ensure that parts of the body and object concepts were listed (Supplementary Table 1).[9]

To extract strict colexifications from the word lists automatically, the data need to be represented in a specific format. The Cross-Linguistic Data Formats (Forkel et al. 2018) offer standards for preparing linguistic data in a reusable and interoperable way. This includes mapping lexical forms to standardised concepts in Concepticon (List et al. 2016; Tjuka et al. 2023). For example, in the word list by Marrison (1967), the Burmese (nucl1310) form *khoṅ* is represented as two separate entries: (1) *khoṅ* 'head' which is mapped to 1256 HEAD and (2) *khoṅ* 'roof' mapped to 769 ROOF. The code provided in a Python 3 (Van Rossum and Drake 2009) script creates the colexifications by aligning the lexical forms in a given word list and checking whether the same form, i.e., *khoṅ*, is mapped to two different concepts, i.e., 1256 HEAD and 769 ROOF.[10] If such a match occurs, a link between the two concepts is drawn and one occurrence of the colexification HEAD-ROOF is counted. Note that only strict colexifications are identified by the automated approach. Instances of loose colexifications such as Mbum (mbum1254) *ŋgàŋ-kpù* 'skin-trunk/tree' for BARK (Urban 2012: 57) are not incorporated. This automated comparison results in a list of body-object colexifications and the number of language varieties in which they occur. The methodological contribution of the study is a newly established workflow which extracts colexifications based on a seed list of a particular subset of concepts and the resulting colexifications are defined as "lexical features".[11] This approach differs from the method used in Lexibank, where the lexical features are predefined by the researcher (List et al. 2022: 8–9).

A prior assessment revealed that 778 body-object colexifications across 931 language varieties were found, but over half of them occurred in only one or two language varieties (588 and 93 varieties, respectively).[12] These singleton colexifications are either due to the structure of the data, because the coverage of concepts varies across languages, or they could be cases of homonomy (List et al. 2013: 2–3). To make the available data comparable, an automated evaluation was carried out to identify the language varieties that had a lexical form for the respective body and object concept for at least 20 out of the 100 most frequent colexifications. Thus, the

9 The data and analysis for the present article are available in a GitHub repository: https://github.com/lexibank/tjukabodyobject/tree/v0.1.0.

10 The code that creates the colexifications uses multiple packages and is available in the GitHub repository: https://github.com/lexibank/tjukabodyobject/blob/v0.1.0/cldfbench_tjukabodyobject.py.

11 Lexibank offers the possibility to define lexical features, similar to the features in the World Atlas of Language Structures (Dryer and Haspelmath 2013). For example, the feature 'Hand and Arm' (Brown 2013) in the category 'Lexicon' is included as the feature 'ArmAndHand' in Lexibank (List et al. 2022: 10). Features can be visualised on a map with the Python package CLDFViz (https://pypi.org/project/cldfviz).

12 The number of body-object colexifications refers to types of colexifications, not tokens.

analysis focuses on body-object colexifications that occur in at least three language varieties. In total, 78 body-object colexifications across 396 language varieties belonging to 79 language families were found. They are combinations between 40 body concepts and 50 object concepts. The full list is presented in Supplementary Table 2.

The results of the analysis are compared with data taken from the Database of Semantic Shifts (Zalizniak et al. 2016).[13] The catalogue presents a large number of semantic shifts attested in the world's languages and includes strict and loose colexifications, i.e., cognates and morphological derivation (Zalizniak 2008: 224). It is important to note that the Database of Semantic Shifts (DatSemShifts) includes some of the same datasets as in CLICS[3] (Rzymski et al. 2020), for example, the Intercontinental Dictionary Series (Borin et al. 2013; Key and Comrie 2015) and the Database of the Languages of New Guinea (Greenhill 2015). This limits the comparability between the results from this study and DatSemShifts. However, since DatSemShifts consists of additional sources and contains loose colexifications, it can still function as a reference point.

4 Results

4.1 Frequency

The 15 most frequent body-object colexifications are shown in Table 3. If a body-object colexification is not attested, the language variety either has a loose colexification, uses two separate terms for the two concepts (dislexification), or the lexical form for at least one of the concepts is not available. This information is crucial because it changes the interpretative power of a given pattern. For example, the colexification SHOULDER BLADE-SPADE is found in eight language varieties but lexical forms for at least one concept are not available for 755 language varieties. Thus, the pattern may be even more frequent if more data are added. Identifying these gaps in the data allows future studies to conduct targeted data collection to improve the quality of the comparison. The full list of frequencies can be found in Supplementary Table 3.

SKIN-LEATHER is the most frequent body-object colexification and is attested in 160 language varieties. The frequency may be higher if data from the 616 language varieties are added. The colexification appears only as a strict colexification in Urban (2012: 679) so it is unlikely that an in-depth analysis would uncover loose colexifications. The Database of Semantic Shifts (DatSemShifts) does not list a semantic shift for SKIN-LEATHER and the colexification is not mentioned in any of the above-

13 See https://datsemshift.ru (accessed 18 February 2023).

Table 3: The 15 most frequent body-object colexifications and the number of language varieties in which a particular colexification is attested ("strict"), not attested ("loose/dislexified")*, or data for one or both concepts are not available ("N/A").

Body	Concept	Object	Concept	Strict	Loose/ Dislexified	N/A
	SKIN		LEATHER	160	155	616
	SKIN		BARK	90	593	248
	TESTICLES		EGG	31	443	457
	HEAD		TOP	24	388	519
	NECK		COLLAR	24	279	628
	BUTTOCKS		BOTTOM	15	260	656
	HEART		FIREWOOD	13	590	328
	SKIN		SHELL	13	323	595
	FINGERNAIL		NAILTOOL	12	299	620
	INTESTINES		SAUSAGE	12	140	779
	LIP		EDGE	12	301	618
	MOUTH		EDGE	12	312	607
	BODY		TREETRUNK	9	302	620
	TENDON		ROOT	9	207	715
	SHOULDERBLADE		SPADE	8	168	755

*With the current analysis, I cannot ensure that a colexification does not occur because it is realised as a loose colexification or the two concepts being dislexified.

discussed studies. The colexification between SKIN and LEATHER is likely due to an intermediate colexification between SKIN (HUMAN) and SKIN (ANIMAL). The distinction between human versus animal skin is not made explicit in most word lists and DatSemShifts lists three realisations for SKIN (ANIMAL) → SKIN (HUMAN) (shift1219).[14] A qualitative study is necessary to determine the exact connections between the concepts.

The second most frequent body-object colexification is SKIN-BARK in 90 language varieties (see also Section 4.2, Figure 4). Here, lexical forms for most language varieties are available except for 248 language varieties. The frequencies illustrate that body-object colexifications with SKIN have the highest number of occurrences in the language sample. Colexifications such as SKIN-BARK and SKIN-SHELL (13 language

14 The shifts can be accessed at https://datsemshift.ru/.

varieties) indicate that the lexical forms for SKIN are used to refer to outer layers or a kind of protective covering. SKIN is also found in less frequent body-object colexifications: SKIN-SKIN (OF FRUIT) (seven language varieties), SKIN-BOOK (four language varieties), and SKIN-LEAF (three language varieties). The colexifications, except SKIN-LEAF, are discussed in Urban (2012: 677–680) but none of them appears in DatSemShifts and they are rarely mentioned in the literature. An areal typological study of their occurrence may reveal interesting loose colexifications.

The third most frequent body-object colexification TESTICLES-EGG occurs in 31 language varieties. This is ten varieties more than in Brown and Witkowski (1983: 603) who used a sample of 81 languages. Urban (2012: 457) identifies 19 languages which have a strict colexification and nine with a loose colexification of TESTICLES-EGG.

Four out of the 15 most frequent body-object colexifications are with spatial terms: HEAD-TOP (24 language varieties), BUTTOCKS-BOTTOM (15 language varieties), LIP-EDGE (12 language varieties), and MOUTH-EDGE (12 language varieties). This supports the finding that body part terms often grammaticalise into spatial markers (Heine 2014: 17–20). The direction from body concept to object concept of these grammaticalisation patterns is supported by the directions found in DatSemShifts (head → top: shift3527, buttocks ↔ bottom: shift0728, lip → edge: shift0523, mouth → edge: shift0483). In Sections 4.2 and 4.3, I discuss these patterns in more detail.

4.2 Areal and global distributions

Most of the areal and global patterns reported in previous studies are based on an analysis of strict and loose body-object colexifications (e.g., Steinberg 2014; Urban 2012; Wilkins 1996). Therefore, the expectation for the present analysis was that few areal patterns occur. The following discussion compares the three areal patterns of strict body-object colexifications previously reported: FOOT-WHEEL (List et al. 2018) and EAR-LEAF, SKIN-BARK (Gast and Koptjevskaja-Tamm 2019). Since more languages are covered in the present study, I examine whether the patterns appear more widespread. In addition, I discuss a seemingly areal pattern with SHOULDER BLADE and the proposed global patterns of TESTICLE-EGG, EYE-SEED, and EYE-FRUIT. I then discuss one global pattern involving ROOT which has not been discussed in previous studies. The analysis does not aspire to be an areal study because the unevenly distributed data and the absence of loose colexifications would distort the results.

The following maps illustrate the patterns (Figures 2–8). The red dots represent language varieties in which the body-object colexification occurs and the yellow dots depict language varieties which either have a loose colexification or dislexify the two

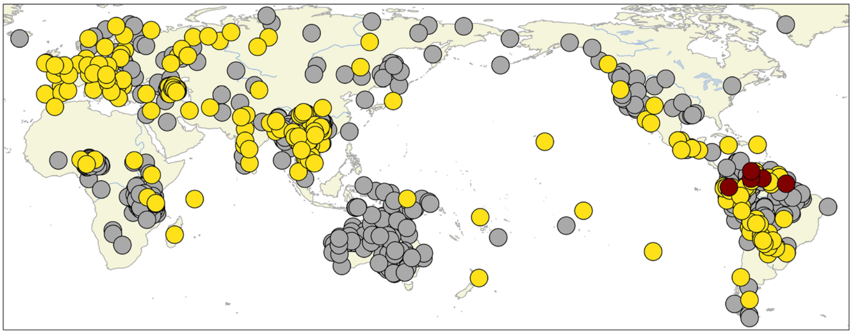

Figure 2: Distribution of languages with the body-object colexification FOOT-WHEEL. Red = colexification attested; yellow = loose colexification or dislexification; grey = data not available.

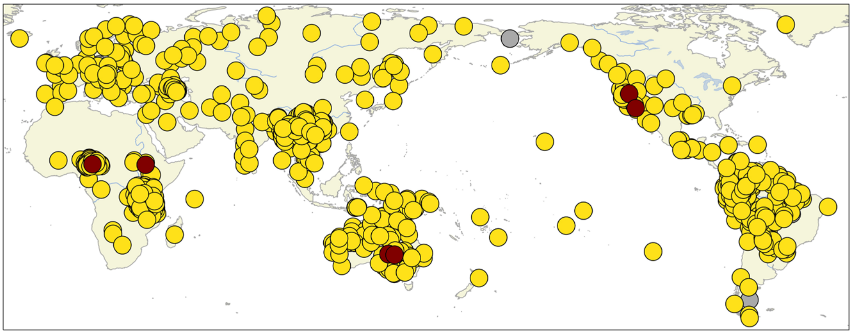

Figure 3: Distribution of languages with the body-object colexification EAR-LEAF. Red = colexification attested; yellow = loose colexification or dislexification; grey = data not available.

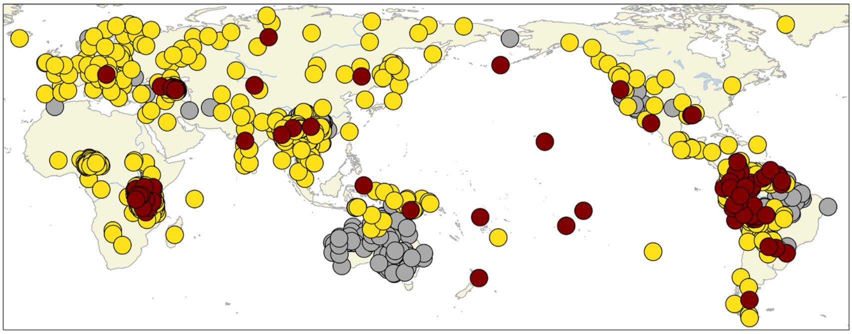

Figure 4: Distribution of languages with the body-object colexification SKIN-BARK. Red = colexification attested; yellow = loose colexification or dislexification; grey = data not available.

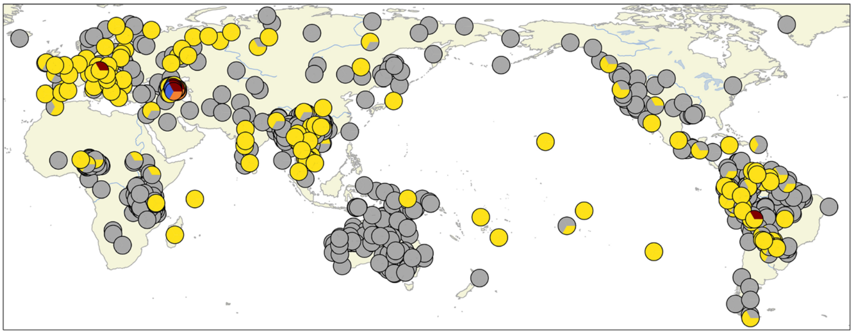

Figure 5: Distribution of languages with body-object colexifications with SHOULDER BLADE. The dots are sliced into three pieces, each representing attested cases of one of the body-object colexifications: SHOULDER BLADE-SPADE (red), SHOULDER BLADE-OAR (blue), and SHOULDER BLADE-PADDLE (orange). Yellow = loose colexification or dislexification; grey = data not available.

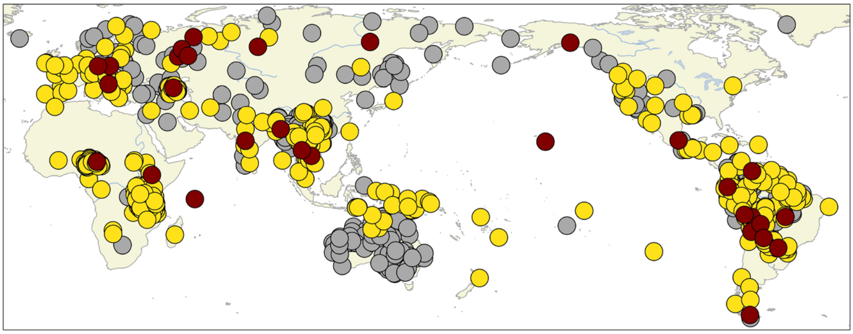

Figure 6: Distribution of languages with the body-object colexification TESTICLES-EGG. Red = colexification attested; yellow = loose colexification or dislexification; grey = data not available.

concepts. The grey dots illustrate the language varieties for which data for at least one concept are not available.

The first pattern FOOT-WHEEL proposed as an areal pattern of South America by List et al. (2018) persisted as an areal pattern in the present language sample (Figure 2).[15] The pattern occurs exclusively in South America in language varieties which are either isolates or belong to different families: Cofán (cofa1242), Puinave (puin1248), Yaruro (pume1238), Wayampi (waya1270), and Ninam (nina1238). Although no other

15 This body-object colexification exists in some African languages, for example, Grebo (greb1257) *bo* 'foot, leg, wheel (of a vehicle), root' (Innes 1967: 12). The present data have gaps in Africa but do not yet confirm that the colexification occurs across African languages in general.

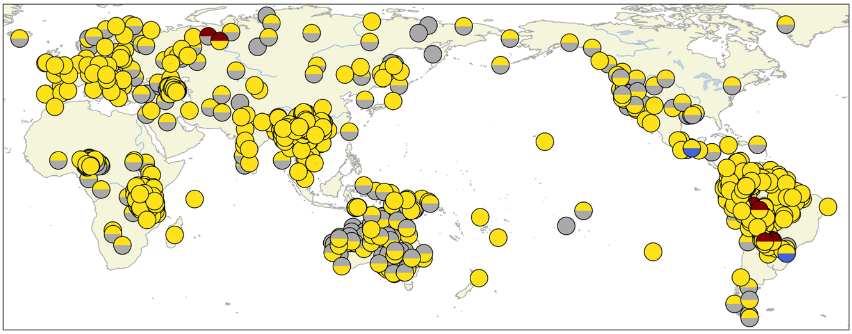

Figure 7: Distribution of languages with the body-object colexification EYE-SEED and EYE-FRUIT. The dots are sliced into two pieces, each half representing attested cases of one of the body-object colexifications: EYE-SEED (top, red) and EYE-FRUIT (bottom, blue). Yellow = loose colexification or dislexification; grey = data not available.

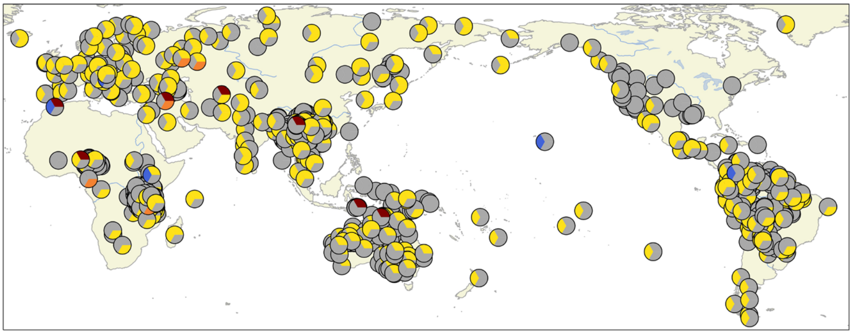

Figure 8: Distribution of languages with body-object colexifications with TENDON. The dots are sliced into three pieces, each representing attested cases of one of the body-object colexifications: TENDON-ROOT (red), BLOOD VESSEL-ROOT (blue), and VEIN-ROOT (orange). Yellow = loose colexification or dislexification; grey = data not available.

body concepts are found to colexify with parts of a car similar to the systematic use in Western Apache (Basso 1967: 472), FOOT-WHEEL may still be more widespread than it appears in the data. Hilpert (2007) found the colexification FOOT-WHEEL in six languages.[16] In DatSemShifts, a shift from FOOT to WHEEL is recorded for three language varieties outside of South America (shift5287): Nuguria (nuku1259), Kapingamarangi

16 Unfortunately, the full list of references and language family affiliations of Hilpert's language sample is not retrievable from the provided web link any more (Footnote 3 in Hilpert 2007: 19) so a comparison of the languages in his sample and the present sample is not possible.

(kapi1249) and Navajo (nava1243). The colexification occurs due to the spatial orientation of the feet and the wheels of a vehicle.

Another areal pattern, found in Eastern Africa, the Americas, and Australia, is EAR-LEAF (Gast and Koptjevskaja-Tamm 2019: 66–68). This finding is supported by the present data although the pattern is less frequent than in the previous study. The authors found a dense meso cluster of 39 languages from 14 families (Gast and Koptjevskaja-Tamm 2019: 66). The areal pattern for EAR-LEAF is found in the ASJP data (Wichmann et al. 2016) instead of CLICS[1] (List et al. 2013) which explains why the pattern is less frequent in the present sample. Figure 3 shows that with the improved data coverage, six language varieties in the areas defined by Gast and Koptjevskaja-Tamm (2019) now emerge for the CLICS data. One realisation in Fula (fula1264) without a direction is listed in DatSemShifts (shift2785). The findings in Steinberg (2014: 281) show a direction from LEAF to EAR in Tzeltal.

The third proposed areal cluster is SKIN-BARK (Figure 4) which occurred in South America and Melanesia in the ASJP data (Gast and Koptjevskaja-Tamm 2019: 68–70). As shown in Table 3, the body-object colexification SKIN-BARK appears across 90 language varieties. In addition to the cluster in South America, a group of languages in Africa shows the pattern. This could be an areal pattern, but due to the sparse data available for African languages, this needs to be further investigated. It is noticeable that ASJP has better coverage for African languages but the data are complementary to CLICS rather than overlapping. To get a better understanding of the linguistic diversity of body-object colexifications in Africa, future studies need to combine the data from CLICS[3] (Rzymski et al. 2020), ASJP (Wichmann et al. 2022), DatSemShifts (Zalizniak et al. 2016), and RefLex (Segerer and Flavier 2022).[17]

The lexico-semantic association between SKIN and BARK is based on a metaphorical transfer (Urban 2011, 2012; Wilkins 1996). Examples of loose colexifications such as *ŋgàŋ-kpù* 'skin-trunk/tree' in Mbum or *hol-chíla* 'tree-skin' in Wappo (wapp1239) meaning BARK are discussed in Urban (2012: 137,140). Although Urban (2011: 6) proposes a scenario in which the direction of a semantic change goes from SKIN to BARK, this direction is only supported by four out of ten realisations in DatSemShifts (shift2326).

Some seemingly areal patterns, such as the body-object colexifications with SHOULDER BLADE shown in Figure 5, are likely due to genealogical relatedness. The colexifications SHOULDER BLADE-SPADE (eight language varieties), SHOULDER BLADE-OAR (six language varieties), and SHOULDER BLADE-PADDLE (six language varieties) mainly occur in Nakh-Dagestanian languages. Although the data for the two concepts are more sparse compared to other body-object colexifications, Nakh-Dagestanian languages tend to colexify different objects with SHOULDER BLADE. The three body-object

---

17 As pointed out by one reviewer, the RefLex database (Segerer and Flavier 2022) lists additional occurrences of the colexification SKIN-BARK in African languages.

colexifications appear in Azerbaijani (nort2697) as well. In addition, the colexification SHOULDER BLADE-SPADE occurs in two other language varieties outside the Nakh-Dagestanian language family: Czech (czec1258) and Cavineña (cavi1250). Cofán (cofa1242) has the colexification SHOULDER BLADE-SHOVEL. Apresjan (1974: 16) discusses the meanings of Polish *lopatka* 'shoulder blade, shovel' as an example of irregular polysemy, i.e., a metaphorical transfer. DatSemShifts lists realisations for a shift from SPADE or SHOVEL to SHOULDER BLADE with eight realisations (shift1918) and a shift from OAR to SHOULDER BLADE with three realisations (shift0112): Tibetan (tibe1272), Armenian (nucl1235) and Azerbaijani. To my knowledge, this pattern has not been described in detail yet and may be a good case for studies in areal semantics.

Three body-object colexifications were proposed as global or widespread patterns in previous studies: TESTICLES-EGG (cf. Brown and Witkowski 1981; Urban 2012; Wilkins 1996), EYE-SEED (Brown and Witkowski 1983; Urban 2012; Steinberg 2014), and EYE-FRUIT (Brown and Witkowski 1983; Urban 2012; Steinberg 2014). The body-object colexification TESTICLES-EGG occurs in 31 language varieties which belong to different language families and are spoken in different geographic areas, as shown in Figure 6. The global distribution of this body-object colexification supports previous findings and demonstrates that it arises independently. DatSemShifts lists 20 realisations, most with a direction from the object concept to the body concept (shift0147). Other less frequently occurring related body-object colexifications include TESTICLES-SEED (five language varieties), TESTICLES-FRUIT (four language varieties), and TESTICLES-BALL (three language varieties) which are discussed in the next section (Section 4.3).

Figure 7 illustrates that only a few language varieties in the sample have either EYE-SEED or EYE-FRUIT as a strict body-object colexification in their lexicon. EYE-SEED appears in seven language varieties and EYE-FRUIT in three language varieties. The body-object colexifications arise in diverse language varieties which indicates an independent development supporting the findings in Brown and Witkowski (1983: 77). The study covers a larger number of languages, but the occurrence of both body-object colexifications does not become more frequent. A small cluster of EYE-SEED arises in South America and it may be more frequent in this area if loose colexifications are considered. This would support the claim that the colexification EYE-SEED is a lexical feature of Mesoamerican languages (Smith-Stark 1982 quoted in Campbell et al. 1986: 553). In addition, Urban (2012: 524) lists *rak-kirik-uˀ* 'wood-eye-NOM' in Pawnee (pawn1254) spoken in North America as a loose colexification for the concept SEED. Another example is the compound *ovo.hae* lit. 'eye.seed' meaning 'eye' in Orokolo (orok1267) (Steinberg 2014: 219). The colexification EYE-SEED is attested in DatSemShifts with two realisations in Hungarian (hung1274) and Malak-Malak[18] (shift4302).

18 No Glottocode for the language name was found in Glottolog Version 4.7 (Hammarström et al. 2022).

Other body-object colexifications show widespread patterns too. Here, I concentrate on one example which has received less attention in previous studies. The body-object colexification TENDON-ROOT (nine language varieties) appears across diverse language varieties. In addition, ROOT is colexified with BLOOD VESSEL (seven language varieties) and VEIN (six language varieties). Figure 8 illustrates that these body-object colexifications emerge in different geographic areas. DatSemShifts attests BLOOD VESSEL-ROOT (shift1020) and TENDON-ROOT (shift3634) in six and two language varieties, respectively. The realisations in the former case show that there are some languages which have a colexification between three of the concepts. In the present data, body-object colexifications of this sort occur in Northern Kurdish (nort2641) which colexifies TENDON-VEIN-ROOT and Tarifiyt Berber (tari1263) which colexifies TENDON-BLOOD VESSEL-ROOT. Again, the data for these concepts are scarcer compared to concepts such as SKIN or BARK because inner body parts are often not featured on word lists used for language comparison. Urban (2012: 521–524) discusses cases of loose colexifications such as Bwe Karen (bwek1238) *θo-kha-wi* 'tree-leg-vein.'

4.3 Cognitive relations

To visualise the connections drawn between body and object concepts, I built a colexification network with Cytoscape (Shannon et al. 2003) in which the nodes represent a concept and the edges (i.e., lines) indicate whether two concepts are colexified (Figure 9). Body concepts are shown in blue (ellipsis shape) and object concepts in orange (rectangle shape). The thickness of the lines represents the

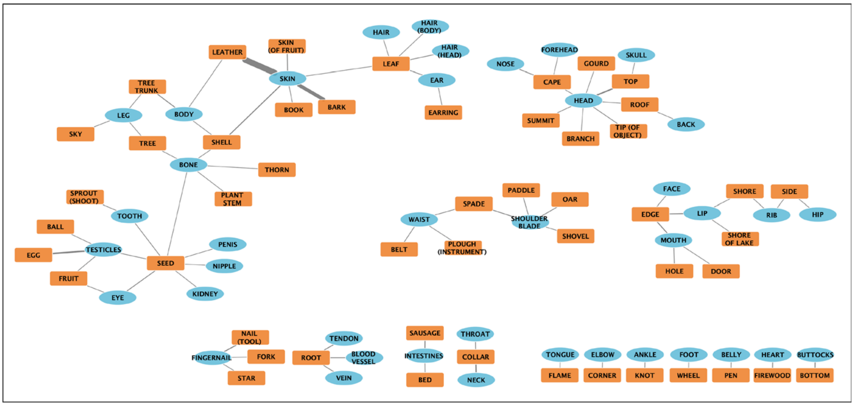

Figure 9: Network of all attested body-object colexifications. Blue ellipsis = body concepts; orange rectangle = object concepts; grey line = colexification attested; thickness = frequency.

number of language varieties in which a body-object colexification is attested, i.e., the frequency (see also Supplementary Table 3).

The network illustrates that there are only a few large components[19] arising from the data including the largest component spanning from SEED to BONE, SHELL, and LEAF. Another dense component is formed with HEAD. More sparse components consist of colexifications with WAIST, SPADE, and SHOULDER BLADE on the one hand and MOUTH, EDGE, and LIP on the other hand. Small components of four to one body-object colexifications are not connected with larger parts of the network. The connections also illustrate that certain body concepts are colexified with multiple object concepts, such as FINGERNAIL which colexifies with NAIL (TOOL), FORK, and STAR. In contrast, TENDON, BLOOD VESSEL, and VEIN all colexify with one object concept, i.e., ROOT. For complete lists of the number of colexifications with each body concept and object concept, see Supplementary Tables 4 and 5.

One of the densest components in the network is the component with HEAD. Figure 10 zooms in on this component and shows that HEAD colexifies with seven different object concepts. The most frequent one is HEAD-TOP (24 language varieties) of which the object concept TOP colexifies with SKULL in three language varieties: Yuwana (yuwa1244), Ngindo (ngin1244), and Kwazá (kwaza1243). Steinberg (2014: 298–303) found mainly examples of semantic changes from a round object to the meaning of SKULL, instead of meaning extensions based on spatial terms. Both body-object colexifications, HEAD-TOP and SKULL-TOP, use the spatial orientation of the head or skull as the top part to establish a cognitive relation. Similarly, HEAD colexifies with ROOF (five language varieties), TIP (OF OBJECT) (five language varieties), SUMMIT (four language

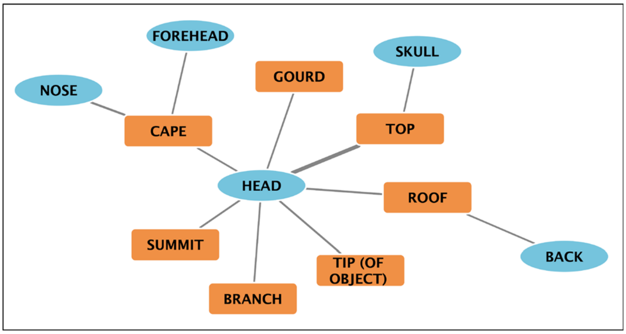

Figure 10: Network with the concept HEAD in the centre. Blue ellipsis = body concepts; orange rectangle = object concepts; grey line = colexification attested; thickness = frequency.

19 A component of a network is defined as a self-contained group of connected nodes that form a part of the network (Newman 2010/2018: 133–134).

varieties), and CAPE (three language varieties). The colexification HEAD-SUMMIT occurs more frequently in DatSemShifts with 23 realisations (shift0515). Semantic shifts between the lexical forms for the two concepts are mostly attested with the direction from HEAD to SUMMIT with one exception in Mongolian (mong1331) where the shift occurred in the reverse order. Furthermore, HEAD colexifies with CAPE which in turn colexifies with NOSE (seven language varieties) and FOREHEAD (three language varieties). While the latter is based on spatial orientation, the former arises due to a similarity in shape between the body part and the object. The colexification NOSE-CAPE is more frequent than the colexification HEAD-CAPE. This finding supports Urban (2012: 479–481) where the lexico-semantic association between NOSE-CAPE is the most frequent. DatSemShifts lists 16 realisations for NOSE-CAPE (shift0490) and one realisation for HEAD-CAPE (shift2229). The direction is from NOSE to CAPE in most cases although a related semantic change emerged from Old Hungarian *orrh* 'summit, ledge' to Hungarian *orr* 'nose' (Steinberg 2014: 274). The colexification NOSE-CAPE emerges independently due to the cognitive relation of similarity in shape, whereas HEAD-CAPE is likely due to genealogical relatedness. The three language varieties with this body-object colexification are Indo-European languages: Irish (iris1253), Italian (ital1282), Romanian (roma1327); and the English word *cape* is derived from Latin *caput* 'head'. This suggests that the body-object colexification is due to a common origin rather than independent innovations. It is noticeable that only one colexification with HEAD is based on its round shape, namely HEAD-GOURD (three language varieties). This finding is in contrast to claims that the roundness of the head leads to frequent extensions of the lexical forms for HEAD (Andersen 1978: 362–364). The low frequency of colexifications between HEAD and round objects in the present sample may be due to the word lists, which often do not include concepts for round objects. Qualitative studies of individual languages show multiple examples of lexical forms for HEAD being used for round objects (Kraska-Szlenk 2019: 138–139).

Another component in which spatial orientation plays a role has the concepts MOUTH and LIP in its centre (Figure 11). Both body concepts colexify with the object concept EDGE which in turn colexifies with FACE. The latter is the least frequent one, occurring in three language varieties, whereas MOUTH-EDGE and LIP-EDGE each occur in 12 language varieties. The body-object colexification MOUTH-EDGE is similarly frequent in Hilpert (2007: 8) where it occurs in 11 languages. In DatSemShifts, both connections are attested with six realisations (mouth → edge: shift0483; lip → edge: shift0523). The direction from MOUTH to EDGE supports the results in Steinberg (2014: 270–271). While MOUTH-EDGE is based on spatial orientation, the other two body-object colexifications with MOUTH are based on the functional perception of the mouth as an opening: MOUTH-HOLE (five language varieties) and MOUTH-DOOR (three language varieties). Both arise in language varieties belonging to different language families and spoken in different geographical areas. They are attested as semantic shifts in

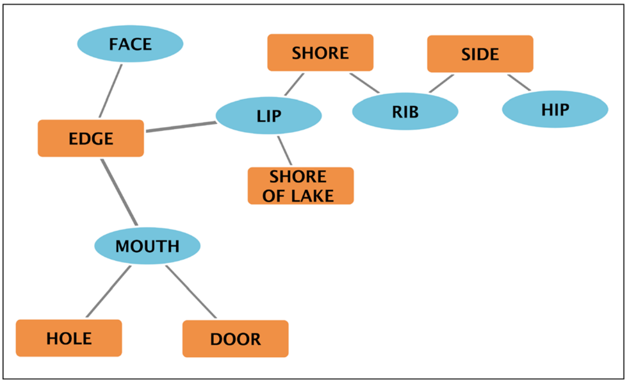

Figure 11: Network with the concept MOUTH and LIP in the centre. Blue ellipsis = body concepts; orange rectangle = object concepts; grey line = colexification attested; thickness = frequency.

DatSemShifts with 26 realisations for MOUTH-HOLE (shift0482) and one realisation for MOUTH-DOOR (shift6236). In the case of MOUTH-DOOR, areal phenomena may also play a role. Avram (2020: 185) lists the combination DOOR + MOUTH meaning 'threshold, doorway' as a potential calque in Creole languages in the Atlantic. In addition, expressions such as *mouth of house* for DOOR are shared across Mesoamerican languages (Smith-Stark 1982 quoted in Campbell et al. 1986: 553). The lexico-semantic association of the mouth with an opening supports the findings that lexical forms for MOUTH are used to refer to the opening of a bottle, as in Lao *paak5* 'mouth (of a bottle, a river)' (Enfield 2006: 193) or Khoekhoegowab (khoe1242) *ǂkhoro-b am!nâ-s* 'bottle-M mouth-F' (Tjuka 2019: 39). This lexico-semantic association is also reflected in the English compound *river mouth*. Integrating loose colexifications would lead to a denser network surrounding the concept MOUTH as shown in Figure 10 in Urban (2012: 322) which includes lexico-semantic associations with CAVE, SPRING, and COAST.

Another interpretation of these body-object colexifications is that MOUTH and LIP are more generally associated with BORDER. For example, the locative prefix *ixi-* 'edge, border' in Lowland Chontal (lowl1260) is used to form landscape terms such as *ixiane'* 'side of a path or road', *ixipana'* 'riverbank', and *ixiñe'maja'* 'seashore' (O'Connor and Kroefges 2008: 300). In addition, Steinberg (2014: 264–265) lists multiple examples in which a lexical form for EDGE developed into a lexical form for LIP. The lexical forms for LIP or MOUTH may have a similar function. Thus, it is not clear whether the body-object colexifications between LIP and SHORE or SHORE OF LAKE are established based on a spatial orientation or a similarity in shape. In comparison to RIB-SHORE, the interpretation based on shape is dominant. However, going further in the network, the body-object colexification RIB-SIDE (three language varieties) is based on the spatial orientation. It is attested with one realisation in DatSemShifts (shift5813) and FACE-SIDE with two

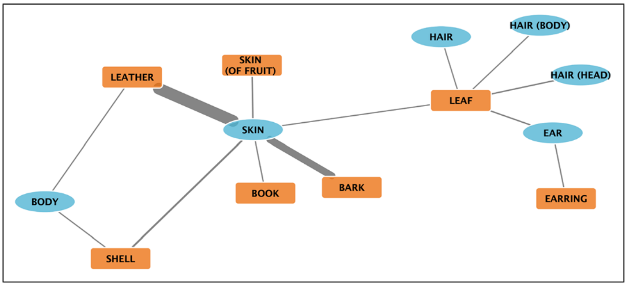

Figure 12: Network with the concept SKIN in the centre. Blue ellipsis = body concepts; orange rectangle = object concepts; grey line = colexification attested; thickness = frequency.

realisations (shift2905). The latter is used in Vietnamese (viet1252) where the sides of a cupboard are denoted with *mặt* 'face' (Tjuka 2023: 249). The colexifications with MOUTH and LIP illustrate that the boundaries between similarity based on function, spatial orientation, and shape are not always straightforward to distinguish. Exploring loose colexifications with other body concepts such as FACE, RIB, and HIP could lead to further insights into the cognitive relations associated with those body parts.

The component with SKIN in Figure 12 demonstrates that this concept frequently colexifies with LEATHER, BARK, and SHELL (see also Table 3, Section 4.1) and with three other concepts: SKIN (OF FRUIT) (seven language varieties), BOOK (four language varieties), LEAF (three language varieties). They are based on a lexico-semantic association of the skin as an outer layer (Urban 2012: 676–680). Another possibility for the interpretation of SKIN-BOOK is the metonymy THE MATERIAL CONSTITUTING AN OBJECT FOR THE OBJECT (Kövecses and Radden 1998: 51) since books were traditionally covered with leather. It is not clear, however, whether this was the case in the region of the four Nakh-Dagestanian languages in which the colexification occurs. The object concepts LEATHER and SHELL colexify with BODY in three language varieties and they are based on an association of shells as the body of animals like turtles. Another interpretation of the colexification between LEATHER and BODY is an intermediate colexification based on metonymy with SKIN. There would be an edge between BODY and SKIN[20] but since the present study focuses on body-object colexifications, this connection is not shown in the current network.

The colexification SKIN-LEAF is the least frequent of the six colexifications with SKIN. There is no reported semantic shift in DatSemShifts and the body-object colexification is not discussed as a loose colexification in previous studies. Although two of the three language varieties with SKIN-LEAF, Mbunga (mbnu1248) and Ngindo (ngin1244), belong to the same language family, the third language variety, Kwazá

20 See https://clics.clld.org/edges/763-1480.

(kwaz1243), is not spoken in close geographic proximity to the other two language varieties. The colexification between SKIN and LEAF may be due to an implicit colexification between SKIN-HUSK since, for example, the husks of corn look like leaves. Urban (2012: 678) lists 11 languages with this colexification. The concept LEAF colexifies with either lexical forms for HAIR or EAR in six language varieties. The concept HAIR is split into multiple sub-concepts in Concepticon including 2648 HAIR (HEAD) and 189 HAIR (BODY). The reason for the split is that there are multiple types of HAIR on the body. If the occurrences of all three HAIR concepts are considered, the colexification HAIR-LEAF is attested for 12 language varieties. Compared to EAR-LEAF, no semantic shift or discussion in the literature is found. The body-object colexification HAIR-LEAF is established based on a part-of relation since both concepts are considered to be part of a larger whole, i.e., the body or the tree. In contrast, EAR-LEAF arises from a similarity based on the shared curved shape.

EAR colexifies with EARRING (four language varieties) but here the connection is due to contiguity which indicates a metonymical mapping. Wilkins (1996: 272) found a shift of EARRING to EAR in three Dravidian languages. The present occurrences are attested in two Hmong-Mien languages, one Pano-Tacanan language, and a Puelche language. This suggests that the pattern is more widespread. Although Steinberg (2014) does not find a semantic change from EAR to EARRING, in Bulgarian (bulg1262) the lexical form for EARLOBE developed from the meaning EARRING (Steinberg 2014: 289). Two other colexifications between body concepts and pieces of clothing appear in the data: NECK-COLLAR (24 language varieties) and THROAT-COLLAR (eight language varieties). Two realisations of the former are attested in DatSemShifts (shift1573). A study examining the use of body part terms for clothing items could lead to interesting cross-linguistic patterns of metonymic extensions.

The concepts LEG, BODY, and BONE form a component (Figure 13). BONE colexifies with five object concepts of which four belong to plants, i.e., THORN, PLANT STEM, SEED, and TREE, all occurring in three language varieties. Except for the colexification BONE-THORN, all other colexifications with BONE are discussed in Urban (2012: 621–622). The colexification BONE-PLANT STEM is listed with one realisation in Yaqui (yaqu1251) in DatSemShifts (shift3900). Trees and other plants evoke a strong association with the human body indicated by frequent body-object colexifications: SKIN-BARK, BODY-TREE TRUNK, and TENDON-ROOT (Table 3, Section 4.1). Another example is LEG-ROOT (Nguyen 2014; Tjuka 2019). While most of the body-object colexifications with parts of the tree are part-of relations (cf. Gentner 1977), for example, HAIR-LEAF, other interpretations are also possible. Schapper (2022) demonstrates that the concept BONE is associated with STRENGTH which is reflected in the firmness of the object concepts THORN and TREE colexifying with BONE.

Figure 14 shows the component with the colexifications between FINGERNAIL and NAIL (TOOL) (12 language varieties), FORK (three language varieties), and STAR (three language varieties) which are based on a similarity in shape. FINGERNAIL-NAIL (TOOL)

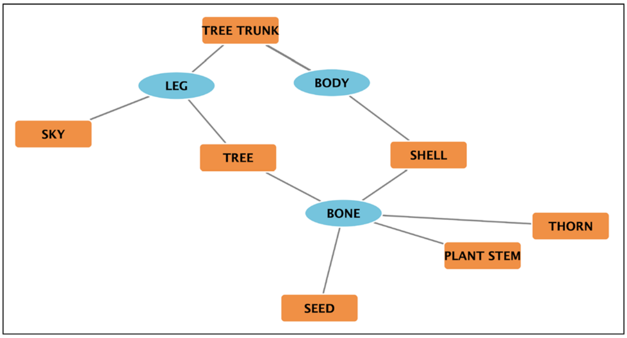

Figure 13: Network with the concept BONE in the centre. Blue ellipsis = body concepts; orange rectangle = object concepts; grey line = colexification attested; thickness = frequency.

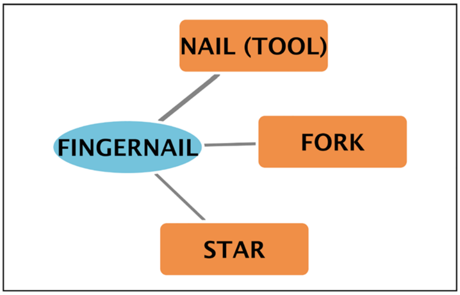

Figure 14: Network with the concept FINGERNAIL in the centre. Blue ellipsis = body concepts; orange rectangle = object concepts; grey line = colexification attested; thickness = frequency.

occurs predominantly in the area of Europe and nine out of the 12 language varieties belong to the Indo-European language family. In DatSemShifts, English and German are listed as realisations of the shift from FINGERNAIL to NAIL (METAL SPIKE) (shift1079). However, there are language varieties such as Polci (polc1243), Khasi (khas1269), and Manchu (manc1252) spoken in other geographic areas which also have the body-object colexification. In comparison, FINGERNAIL-FORK and FINGERNAIL-STAR appear in language varieties of the same language family: Nakh-Dagestanian and Tungusic, respectively. This indicates that genealogical relatedness is a stronger factor than a general cognitive relation based on shape.

The above discussion has focused mainly on components with body concepts at their centre. There are however object concepts that colexify with multiple body concepts. Examples are the colexifications with SEED shown in Figure 15. The object concept colexifies with EYE (seven language varieties), TESTICLES (five language varieties), TOOTH (five language varieties), KIDNEY (three language varieties), NIPPLE (three language varieties), BONE (three language varieties), and PENIS (three language varieties). The colexifications EYE-SEED, TOOTH-SEED, and KIDNEY-SEED are based on a similarity in terms of the round, small shape of seeds. Lexical forms for SEED are also used for

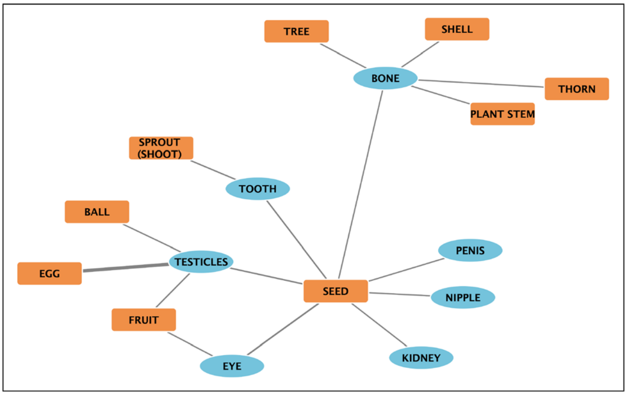

Figure 15: Network with the concept SEED in the centre. Blue ellipsis = body concepts; orange rectangle = object concepts; grey line = colexification attested; thickness = frequency.

TESTICLES (Urban 2012; Wilkins 1996; and four realisations in DatSemShifts: shift4416). While previous studies indicate that the cognitive relation is based on similarity in shape, another interpretation is the association of seed as an offspring so that the relation is rather due to a PRODUCER-FOR-PRODUCT metonymy (Kövecses and Radden 1998: 57) with an implicit metaphorical mapping in between. This interpretation is supported by other reproductive organs, i.e., PENIS and NIPPLE, which colexify with SEED. The use of an object concept rather than an anatomical term for a body part could also be due to taboo conventions. The body concept TESTICLES is prone to be referred to with lexical forms for objects (Urban 2012: 689–691; Wilkins 1996: 273). However, the pattern needs to be further investigated with additional data on the socio-linguistic context in which these body-object colexifications occur. Furthermore, in light of the colexification BONE-SEED, the body-object colexification TOOTH-SEED can be interpreted as based on the firmness of seeds.

4.4 Vision and touch perception

The previous observations were based on a qualitative analysis. In this section, I present a case study which examines ratings on how body and object concepts are perceived visually and haptically. The ratings are taken from the Lancaster Sensorimotor Norms (Lynott et al. 2020), which provide ratings on six perceptual modalities (touch, hearing, smell, taste, vision, and interoception) and other variables for 40,000 English words. English-speaking participants rated the words on a scale from 0 (*not experienced at all*) to 5 (*experienced greatly*) (Lynott et al. 2020: 1275). The words with their associated mean ratings were mapped to the Concepticon and integrated into the

Cross-Linguistic Database of Norms, Ratings, and Relations for Words and Concepts (NoRaRe, Tjuka et al. 2022).[21] From NoRaRe, I extracted the ratings for the body and object concepts that occurred in a given body-object colexification. No ratings were available for the concepts HAIR (BODY), HAIR (HEAD), NAIL (TOOL), SHORE OF LAKE, SPROUT (SHOOT), and SKIN (OF FRUIT),[22] so the number was reduced to 72 body-object colexifications.

The analysis of the interaction between the ratings for visual and haptic perception across the body-object colexifications was performed in R Version 4.3.1 (R Core Team 2022). I used a Bayesian linear regression model included in the *brms* package Version 2.20.4 (Bürkner 2017) to analyse the interaction between perception type (vision versus touch) and used the *posterior* package Version 1.4.1 (Bürkner et al. 2023) to draw samples from the posterior distribution.[23]

On average, body and object concepts had higher ratings for visual perception (body concepts: m = 3.74; object concepts: m = 3.99) compared to haptic perception (body concepts: m = 2.99; object concepts: m = 2.64). Based on an analysis of the Pearson coefficient (Pearson 1895), there was no significant correlation between the mean ratings for vision and touch of the body and object concepts (vision: r = 0.00, p-value = 0.99; touch: r = 0.12, p-value = 0.29).[24]

By integrating perception type as varying residuals and drawing samples from the posterior distribution, I analysed the variation in the standard deviations between vision and touch (vision: sd = 1.81; touch: sd = 2.06). The results indicate that body and object concepts align more closely in their visual perception compared to their haptic perception. The interactions of the mean ratings across body-object colexifications with the body concepts BONE, HEAD, LEG, SHOULDER BLADE, SKIN, and TESTICLES in Figure 16 show that the perception of visual and haptic features varies.

The differences between the alignment of visual and haptic perception in a given body-object colexification are stronger in some cases than in others. For the body concepts BONE, HEAD, and LEG, individual body-object colexifications differ significantly in their vision and touch mean ratings, for instance, BONE-PLANT STEM, HEAD-ROOF, and LEG-SKY. These body-object colexifications do not occur in a wide range of languages. In contrast, the body-object colexifications with SKIN and TESTICLES align closely in their mean ratings across both sensory domains and they occur across diverse languages. This suggests that the more similar ratings are for a body concept and an object

21 See https://norare.clld.org.

22 Note that ratings on psycholinguistics variables do not differentiate the meanings of the words that are presented to participants. Tjuka et al. (2022) discuss the issue of mapping such word lists to Concepticon.

23 The data and script are available in the GitHub repository: https://github.com/lexibank/tjukabodyobject/tree/v0.1.0/scripts.

24 Note that the ratings of the sensory modalities can correlate with each other. This can lead to issues of multicollinearity.

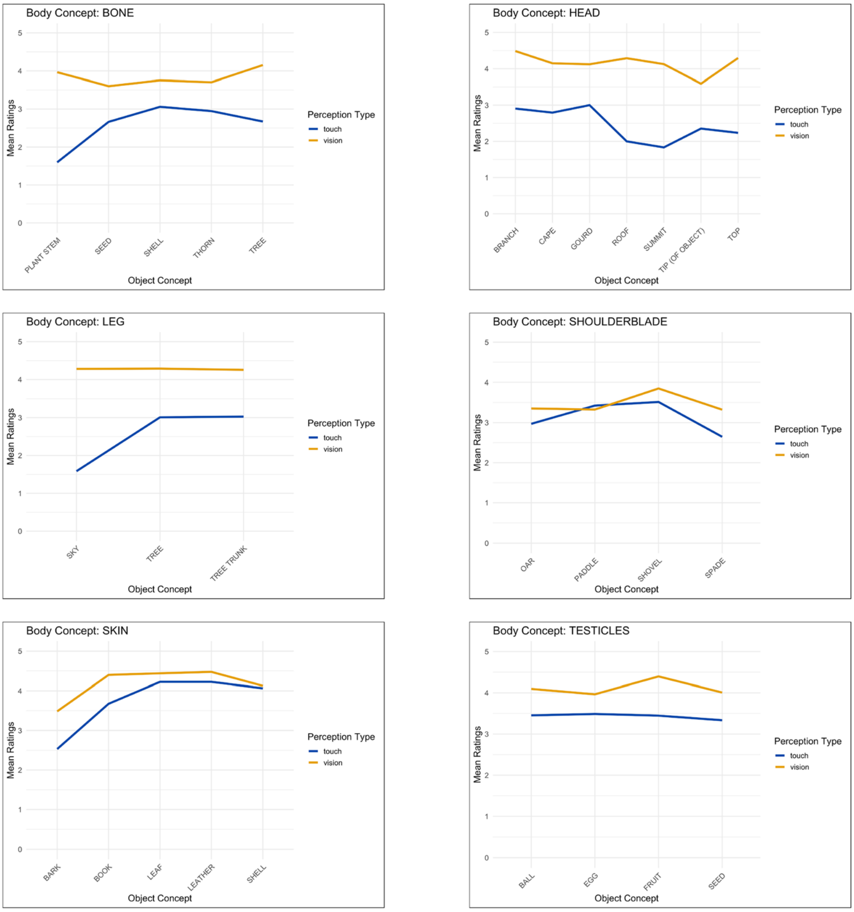

Figure 16: Interaction plots of mean ratings for a subset of body-object colexifications. The plots represent the relationship between perception types (vision and touch). The *x*-axis shows the object concepts and the *y*-axis mean ratings. Ratings on vision are displayed in orange, and ratings on touch in blue.

concept, the more common a colexification between the two concepts is in the different languages. Outliers are the body-object colexifications with SHOULDER BLADE which align in their perception ratings but do not occur in a large number of languages.

The findings of the case study have important implications for explanations involving similarity and salience. A body-object colexification can arise due to a similarity in the perception of vision but in the case of touch, the perception can be asymmetrical. However, the model does not factor in whether the body or object concept is rated higher for visual/haptic perception so no implication about directionality can be drawn. The case study provides a first overview of the interaction between visual and haptic perception in body-object colexifications. It draws from an existing set of data but in a future study, additional data on all 78 body-object colexifications need to be collected. In addition, ratings from other languages need to be included to draw conclusions about cross-linguistic differences in perception.

4.5 Coincidence and cultural variation

Multiple cognitive relations can lead to body-object colexifications. In certain cases, additional data are necessary to disentangle the influence of cognitive relations from historical processes and the socio-linguistic context in which the languages are spoken. The latter also leads to the discovery of cultural variations that seem coincidental at first sight. Here, I focus on the body-object colexifications HEART-FIREWOOD, BACK-ROOF, and HEAD-BRANCH. Since the present analysis examines body-object colexifications which occur in at least three language varieties, purely coincidental cases are not found.

One frequent body-object colexification is striking in comparison to other frequent colexifications: HEART-FIREWOOD found in 13 language varieties which all belong to the Sino-Tibetan language family (Table 3, Section 4.1). The colexification is attested in two studies on dialects of Bai (baii1251) spoken in China by over one million speakers (Allen 2007; Wang 2004). The colexification is not further discussed in these studies so one can only speculate how the connection between HEART and FIREWOOD arose. It might be a metaphorical mapping due to HEART being the seat of emotion in Chinese culture (cf. Yu 2002). Associations between FIRE and PASSION are found in English expressions such as *There was a spark between us*.[25] A qualitative study of conceptual metaphors in Bai and other languages is necessary to give a conclusive answer to whether this body-object colexification represents a cultural variation or a coincidence.

Another interesting case is BACK-ROOF which occurs in three language varieties in South America: Mocoví (moco1246), Toba (1269), and Siona (sion1247). This body-object colexification is likely due to language contact or a general preference. South American languages tend to use body part terms for objects systematically. Levinson (1994) demonstrates that the language Tzeltal has a sophisticated system for using body part terms for object parts according to their shape. Similarly, Tilbe (2017: 59) shows that Tzeltal and Zapotec speakers refer to objects more frequently with body part terms than English speakers. In the case of BACK-ROOF, a zoomorphic orientation instead of an anthropomorphic orientation is used. Thus, the back of an animal rather than a human gives rise to this body-object colexification. This supports the findings by Bowers (2022: 101) that in Mixtepec-Mixtec (mixt1425), the compound *titsi*

25 Another example is the Swahili (swah1253) expression *moyo wa mti*, which means ‘heart of the tree’ for the inner part of a tree. One reviewer suggested a possible extension based on this association.

meta lit. 'belly table' refers to the underside of a table. A zoomorphic orientation is used here too.

The body-object colexification HEAD-BRANCH first looks coincidental. From a European viewpoint, the branches would intuitively be the arms of the tree. DatSemShifts lists HAND/ARM-BRANCH with nine realisations (shift1206). However, multiple branches on top of the tree constitute the hair in personifications of trees in animated films or fantasy stories. The metonymy between HEAD-HAIR may trigger the body-object colexification between HEAD-BRANCH. This body-object colexification appears in three language varieties belonging to the Tupian language family: Old Guaraní (oldp1258), Araweté (araw1273), and Xeta (xeta1241). It would be interesting to contrast the body-object colexification HEAD-BRANCH with HAND/ARM-BRANCH (Gaby 2006: 218; Urban 2012: 434–437) and LEG-BRANCH (Urban 2012: 434–437) to see what factors influence the colexification of either one of the body concepts with a tree branch.

5 Conclusions

This article has presented the first systematic study on body-object colexifications across a large number of languages. I analysed 78 body-object colexifications occurring across 396 language varieties in terms of their frequency, distribution, cognitive relations, and cultural variations. The results show that the phenomenon is complex and that there is a great deal of linguistic diversity. While some widespread body-object colexifications such as SKIN-BARK or TESTICLES-EGG exist, most body-object colexifications occur in a small number of language varieties. By using an increased amount of data, the study revealed body-object colexifications not discussed in previous literature including SKIN-LEATHER, FINGERNAIL-NAIL (TOOL), or HEART-FIREWOOD. The list of body-object colexifications presented here can be used as a starting point for further investigations.

The methodological contributions of the study are twofold. First, the study presents a new workflow that uses 36 datasets from Lexibank (List et al. 2022) to identify colexifications automatically. By adapting the methods presented in List et al. (2022), the colexifications were extracted based on a seed list of 134 body concepts and 650 object concepts. Instead of predefining the colexifications as lexical features, they were selected based on their occurrence. Second, analysing the occurrence of different body-object colexifications in network visualisations is essential. The networks can be further analysed in the semantic maps framework to establish generalisations based on synchronic and diachronic data (cf. François 2021; Georgakopoulos and Polis 2022).

The theoretical implications of the study are important in terms of how body-object colexifications are categorised into theoretical constructs. Examples of body-object colexifications are usually analysed as metaphor, metonymy, or meronomy (except for Levinson 1994 who argues for a non-metaphorical interpretation). Although these categories still have merit, I argue that it is crucial to consider body-object colexifications on a scale of literal versus figurative similarity. Instead of finding an all-encompassing definition, my approach involved analysing speakers' judgements of how individual concepts are perceived and comparing ratings on vision and touch. The case study showed that close alignment of ratings on vision and touch is related to literal similarity and higher frequency, whereas divergence in alignment is related to figurative similarity and low frequency across languages. These results represent a concrete prediction that can be tested with more data. By including a large number of body and object concepts and psycholinguistic measures, the study contributes to evaluating the subjective interpretation of salience and similarity used in previous studies with a quantitative approach.

However, the large-scale approach has limitations. In particular, areal patterns reflected in the occurrence of loose colexifications that are based on a similar lexico-semantic association fall through the coarse grid of the automated method. Detailed analyses in areal semantics offer a more nuanced view of particular lexical patterns (e.g., Schapper et al. 2016; Schapper 2022; Urban 2022). While lexical databases can provide an overarching picture they should be supplemented with additional evidence from qualitative studies. For this purpose, the results were compared with attested realisations in the Database of Semantic Shifts (Zalizniak et al. 2016) and examples from previous studies (e.g., Brown and Witkowski 1981, 1983; Hilpert 2007; Steinberg 2014; Urban 2012; Wilkins 1996). In the future, the strict colexifications can be supplemented with loose colexifications by using the methods presented in List (2023). Furthermore, preliminarily identified areal patterns such as the colexifications with SHOULDER BLADE could be analysed in more detail.

Linguistic variation is a challenge from a methodological and theoretical point of view. The study addressed these challenges by qualitatively and quantitatively investigating body-object colexifications and by suggesting new avenues for further research.

Abbreviations

ASJP	Automated Similarity Judgment Program (Wichmann et al. 2022)
CLDF	Cross-Linguistic Data Formats (Forkel et al. 2018)
CLICS	Cross-Linguistic Database of Colexifications (Rzymski et al. 2020)
DatSemShifts	Database of Semantic Shifts (Zalizniak et al. 2016)
F	feminine
M	masculine
NOM	nominative

Acknowledgements: I would like to thank Kilu von Prince for sparking my interest in this research topic. Special thanks to Maïa Ponsonnet for her feedback on an earlier version of the manuscript. A big thank you to Johannes Englisch and Johann-Mattis List for helping me to implement this study into the Lexibank framework. As the author, I am solely responsible for any mistakes. This work was supported by a fellowship from the International Max Planck Research School (IMPRS) for the Science of Human History.

Data availability: The data that support the findings of this study are openly available on GitHub at https://github.com/lexibank/tjukabodyobject/tree/v0.1.0.

References

Allen, Bryan. 2007. *Bai dialect survey*. Dallas: SIL International. Available at: https://www.sil.org/resources/archives/9121.

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: Stanford University Press.

Apresjan, Jurij D. 1974. Regular polysemy. *Linguistics* 12(142). 5–32.

Avram, Andrei A. 2020. Idiomatic calques. *B.A.S. British and American Studies* 26(26). 181–199.

Bakker, Dik. 2010. Language sampling. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 100–128. Oxford: Oxford University Press.

Basso, Keith H. 1967. Semantic aspects of linguistic acculturation. *American Anthropologist* 69(5). 471–477.

Behrens, Leila. 2002. Structuring of word meaning II. In D. Alan Cruse, Franz Hundsnurscher, Michael Job & Peter Rolf Lutzeier (eds.), *Lexikologie/Lexicology: 1. Halbband*, 319–337. Berlin: Walter de Gruyter.

Blank, Andreas & Peter Koch (eds.). 1999. *Historical semantics and cognition*. Berlin: De Gruyter Mouton.

Blust, Robert. 2011. 'Eye of the day': A response to urban (2010). *Oceanic Linguistics* 50(2). 524–535.

Borin, Lars, Bernard Comrie & Anju Saxena. 2013. The intercontinental dictionary series. In Lars Borin & Anju Saxena (eds.), *Approaches to measuring linguistic differences*, 285–302. Berlin: De Gruyter Mouton.

Bowers, Jack. 2022. Pathways and patterns of metaphor and metonymy in Mixtepec-Mixtec body-part terms. In Roberto Zariquiey & Pilar M. Valenzuela (eds.), *The grammar of body-part expressions*, 91–125. Oxford: Oxford University Press.

Brenzinger, Matthias & Iwona Kraska-Szlenk (eds.). 2014. *The body in language – Comparative studies of linguistic embodiment*. Leiden: Brill.

Brown, Cecil H. 1976. General principles of human anatomical partonomy and speculations on the growth of partonomic nomenclature. *American Ethnologist* 3(3). 400–424.

Brown, Cecil H. 2013. Hand and arm. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/129.

Brown, Cecil H. & Stanley R. Witkowski. 1981. Figurative language in a Universalist perspective. *American Ethnologist* 8(3). 596–615.

Brown, Cecil H. & Stanley R. Witkowski. 1983. Polysemy, lexical change and cultural importance. *Man* 18(1). 72–89.

Buck, Carl Darling. 1949/1988. *A dictionary of selected synonyms in the principal Indo-European languages*. Chicago: University of Chicago Press.
Bürkner, Paul-Christian. 2017. Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28.
Bürkner, Paul-Christian, Jonah Gabry, Matthew Kay & Aki Vehtari. 2023. Posterior: Tools for working with posterior distributions. Availbale at: https://mc-stan.org/posterior/.
Cablitz, Gabriele H. 2008. When "what" is "where": A linguistic analysis of landscape terms, place names and body part terms in Marquesan. *Language Sciences* 30(2). 200–226.
Campbell, Lyle, Terrence Kaufman & Thomas C. Smith-Stark. 1986. Meso-America as a linguistic area. *Language* 62(3). 530–570.
Cruse, D. Alan. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.
Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: https://wals.info/.
Enfield, Nicholas J. 2006. Lao body part terms. *Language Sciences* 28(2). 181–200.
Forkel, Robert, Johann-Mattis List, Simon J. Greenhill, Christoph Rzymski, Sebastian Bank, Michael Cysouw, Harald Hammarström, Martin Haspelmath, Gereon A. Kaiping & Russell D. Gray. 2018. Cross-linguistic data formats, advancing data sharing and re-use in comparative linguistics. *Scientific Data* 5(1). 1–10.
François, Alexandre. 2008. Semantic maps and the typology of colexification. In Martine Vanhove (ed.), *From polysemy to semantic change*, 163–215. Amsterdam: John Benjamins.
François, Alexandre. 2021. Lexical tectonics. *Zeitschrift für Sprachwissenschaft* 41(1). 89–123.
Gaby, Alice R. 2006. The Thaayorre 'true man'. *Language Sciences* 28(2). 201–220.
Gast, Volker & Maria Koptjevskaja-Tamm. 2019. The areal factor in lexical typology. In Daniël Van Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of linguistic variation*, 43–82. Berlin: Walter de Gruyter.
Geeraerts, Dirk. 1993. Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics* 4(3). 223–272.
Geeraerts, Dirk. 2010. *Theories of lexical semantics*. Oxford: Oxford University Press.
Geeraerts, Dirk. 2015. How words and vocabularies change. In John R. Taylor (ed.), *The Oxford handbook of the word*, 416–430. Oxford: Oxford University Press.
Gentner, Dedre. 1977. If a tree had a knee, where would it be? Children's performance on simple spatial metaphors. *Papers and Reports on Child Language Development* 13. 157–164.
Georgakopoulos, Thanasis & Stéphane Polis. 2022. New avenues and challenges in semantic map research. *Zeitschrift für Sprachwissenschaft* 41(1). 1–30.
Goossens, Louis. 1990. Metaphtonymy: The interaction of metaphor and metonymy in expressions for linguistic action. *Cognitive Linguistics* 1(3). 323–342.
Greenhill, Simon J. 2015. TransNewGuinea.org: An online database of New Guinea languages. *PLoS One* 10(10). 1–17.
Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2022. *Glottolog 4.7*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: https://glottolog.org.
Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.
Heine, Bernd. 2014. The body in language: Observations from grammaticalization. In Matthias Brenzinger & Iwona Kraska-Szlenk (eds.), *The body in language – Comparative studies of linguistic embodiment*, 11–32. Leiden: Brill.
Hilpert, Martin. 2007. Chained metonymies in lexicon and grammar. In Günter Radden (ed.), *Aspects of meaning construction*, 77–98. Amsterdam: John Benjamins.

Hock, Hans Henrich & Brian D. Joseph. 1996/2019. *Language history, language change, and language relationship*, 3rd edn. Berlin: De Gruyter.
Huisman, John L. A., Roeland van Hout & Asifa Majid. 2021. Patterns of semantic variation differ across body parts. *Cognitive Linguistics* 32(3). 455–486.
Innes, Gordon. 1967. *A Grebo-English dictionary*. London: Cambridge University Press.
Key, Mary Ritchie & Bernard Comrie. 2015. *The intercontinental dictionary series (Version 4.1)*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: https://ids.clld.org.
Key, Mary Ritchie & Bernard Comrie. 2023. *The intercontinental dictionary series (Version 4.3)*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: http://ids.clld.org.
Koch, Peter. 2001. Lexical typology from a cognitive and linguistic point of view. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible. (eds.), *Language typology and language universals*, vol. 2, 1142–1178. Berlin: De Gruyter Mouton.
Koch, Peter. 2008. Cognitive onomasiology and lexical change: Around the eye. In Martine Vanhove (ed.), *From polysemy to semantic change*, 107–137. Amsterdam: John Benjamins.
Koch, Peter. 2011. The pervasiveness of contiguity and metonymy in semantic change. In Kathryn Allan & Justyna A. Robinson (eds.), *Current methods in historical semantics*, 259–312. Berlin: De Gruyter Mouton.
Koch, Peter. 2016. Meaning change and semantic shifts. In Päivi Juvonen & Maria Koptjevskaja-Tamm (eds.), *The lexical typology of semantic shifts*, 21–66. Berlin: De Gruyter Mouton.
Koptjevskaja-Tamm, Maria & Henrik Liljegren. 2017. Semantic patterns from an areal perspective. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 204–236. Cambridge: Cambridge University Press.
Koptjevskaja-Tamm, Maria, Ekaterina Rakhilina & Martine Vanhove. 2015. The semantics of lexical typology. In Nick Riemer (ed.), *The Routledge handbook of semantics*, 434–454. London: Routledge.
Koptjevskaja-Tamm, Maria, Martine Vanhove & Peter Koch. 2007. Typological approaches to lexical semantics. *Linguistic Typology* 11(1). 159–185.
Kövecses, Zoltán & Günter Radden. 1998. Metonymy: Developing a cognitive linguistic view. *Cognitive Linguistics* 9(1). 37–78.
Kraska-Szlenk, Iwona. 2014. Semantic extensions of body part terms. *Language Sciences* 44. 15–39.
Kraska-Szlenk, Iwona. 2019. Metonymic extensions of the body part 'head' in mental and social domains. In Iwona Kraska-Szlenk (ed.), *Embodiment in cross-linguistic studies: The 'head'*, 136–154. Leiden: Brill.
Kraska-Szlenk, Iwona (ed.). 2020. *Body part terms in conceptualization and language usage*. Amsterdam: John Benjamins.
Lakoff, George. 1987/1990. *Women, fire, and dangerous things*. Chicago: University of Chicago Press.
Lakoff, George & Mark Johnson. 1980/2003. *Metaphors we live by*. Chicago: University of Chicago Press.
Lakoff, George & Mark Johnson. 1999. *Philosophy in the flesh*. New York: Basic Books.
Lehrer, Adrienne. 1990. Polysemy, conventionality, and the structure of the lexicon. *Cognitive Linguistics* 1(2). 207–246.
Levinson, Stephen C. 1994. Vision, shape, and linguistic description: Tzeltal body-part terminology and object description. *Linguistics* 32(4–5). 791–855.
List, Johann-Mattis. 2023. Inference of partial colexifications from multilingual wordlists. https://doi.org/10.48550/arXiv.2302.00739.
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 Odijk & Stelios Piperidis (eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, 2393–2400. Portorož, Slovenia: European Language Resources Association. Available at: https://aclanthology.org/L16-1379/.

List, Johann-Mattis, Robert Forkel, Simon J. Greenhill, Christoph Rzymski, Johannes Englisch & Russell D. Gray. 2022. Lexibank, a public repository of standardized wordlists with computed phonological and lexical features. *Scientific Data* 9(1). 316.

List, Johann-Mattis, Simon J. Greenhill, Cormac Anderson, Thomas Mayer, Tiago Tresoldi & 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.

List, Johann-Mattis, Christoph Rzymski, Simon J. Greenhill, Nathanael Schweikhard, Kristina Pianykh, Annika Tjuka, Carolin Hundt & Robert Forkel. 2021. *Concepticon (Version 2.5.0)*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: https://concepticon.clld.org.

List, Johann-Mattis & Anselm Terhalle & Matthias Urban. 2013. Using network approaches to enhance the analysis of cross-linguistic polysemies. In Alexander Koller & Katrin Erk (eds.), *Proceedings of the 10th International Conference on Computational semantics*, 347–353. Potsdam, Germany: Association for Computational Linguistics. Available at: https://www.aclweb.org/anthology/W13-0208.

List, Mattis. 2014. *Sequence comparison in historical linguistics*. Düsseldorf: Düsseldorf University Press dissertation.

Lynott, Dermot, Louise Connell, Marc Brysbaert, James Brand & James Carney. 2020. The Lancaster Sensorimotor Norms. *Behavior Research Methods* 52. 1271–1291.

MacLaury, Robert E. 1989. Zapotec body-part locatives. *International Journal of American Linguistics* 55(2). 119–154.

Majid, Asifa. 2010. Words for parts of the body. In Barbara C. Malt & Phillip Wolff (eds.), *Words and the mind*, 58–71. Oxford: Oxford University Press.

Majid, Asifa, Nicholas J. Enfield & Miriam van Staden. 2006. Parts of the body (special issue). *Language Sciences* 28(2–3). 137–360.

Marrison, Geoffrey Edward. 1967. *The classification of the Naga languages of North-East India*. London: School of African and Oriental Sciences.

Morrison, Julie Bauer & Barbara Tversky. 2005. Bodies and their parts. *Memory & Cognition* 33(4). 696–709.

Newman, Mark. 2010/2018. *Networks: An introduction*, 2nd edn. Oxford: Oxford University Press.

Nguyen, Van Hai. 2014. On the culturology language of human body parts in Vietnamese and English. *International Journal of Language and Linguistics* 2(6). 368–373.

O'Connor, Loretta & Peter C. Kroefges. 2008. The land remembers. *Language Sciences* 30(2). 291–315.

Pearson, Karl. 1895. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London* 58(347–352), 240–242.

R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna. Available at: https://www.R-project.org/.

Riemer, Nick. 2002. Remetonymizing metaphor. *Cognitive Linguistics* 12(4). 379–401.

Rzymski, 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 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 & Johann-Mattis List. 2020. The database of cross-linguistic colexifications, reproducible analysis of cross-linguistic polysemies. *Scientific Data* 7(1). 1–12.

Schapper, Antoinette. 2022. Baring the bones: The lexico-semantic association of bone with strength in Melanesia and the study of colexification. *Linguistic Typology* 26(2). 313–347.

Schapper, Antoinette, Lila San Roque & Rachel Hendery. 2016. Tree, firewood and fire in the languages of Sahul. In Päivi Juvonen & Maria Koptjevskaja-Tamm (eds.), *The lexical typology of semantic shifts*, 355–422. Berlin: Walter de Gruyter.

Segerer, Guillaume & Sébastien Flavier. 2022. RefLex: Reference lexicon of Africa (version 2.0). Paris, Lyon: Centre national de la recherche scientifique. Available at: http://reflex.cnrs.fr/.
Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski & Trey Ideker. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* 13(11). 2498–2504.
Smith-Stark, Thomas C. 1994. Mesoamerican calques. In Carolyn J. Mackay & Verónica Vázquez (eds.), *Investigaciones lingüísticas en Mesoamérica*, 15–50. México: Universidad Nacional Autónoma de México.
Souag, Lameen. 2022. How a West African language becomes North African, and vice versa. *Linguistic Typology* 26(2). 283–312.
Steinberg, Reinhild. 2014. *Lexikalischer Wandel und Polygenese: Der Konzeptbereich KOPF*. Tübingen: Stauffenburg-Verlag Narr.
Swadesh, Morris. 1952. Lexico-statistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96(4). 452–463.
Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21(2). 121–137.
Tadmor, Uri. 2009. Loanwords in the world's languages. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages*, 55–75. Berlin: De Gruyter Mouton.
Tilbe, Timothy James. 2017. *Parts and wholes in Mesoamerican language and cognition*. Buffalo: State University of New York at Buffalo Dissertation.
Tjuka, Annika. 2019. *Body-part metaphors as a window to cognition: A cross-linguistic study of object and landscape terms*. Berlin Humboldt-Universität zu Berlin MA thesis.
Tjuka, Annika. 2020. A list of 171 body part concepts. *Computer-Assisted Language Comparison in Practice* 3(10). 1–3. https://calc.hypotheses.org/2512.
Tjuka, Annika. 2022. A concept list for the study of semantic extensions from body to objects. *Computer-Assisted Language Comparison in Practice* 5(4). 1–6. https://calc.hypotheses.org/3840.
Tjuka, Annika. 2023. Body part extensions with mặt 'face' in Vietnamese. In Kelsie E. Pattillo & Małgorzata Waśniewska (eds.), *Embodiment in cross-linguistic studies: The 'face*, 237–255. Leiden: Brill.
Tjuka, Annika, Robert Forkel & Johann-Mattis List. 2022. Linking norms, ratings, and relations of words and concepts across multiple language varieties. *Behavior Research Methods* 54. 864–884. https://norare.clld.org/.
Tjuka, Annika, Robert Forkel & Johann-Mattis List. 2023. Curating and extending data for language comparison in Concepticon and NoRaRe. *Open Research Europe* 2(141). 1–13.
Traugott, Elizabeth Closs & Richard B. Dasher. 2001. *Regularity in semantic change*. Cambridge: Cambridge University Press.
Tuggy, David. 1993. Ambiguity, polysemy, and vagueness. *Cognitive Linguistics* 4(3). 273–290.
Tversky, Barbara & Kathleen Hemenway. 1984. Objects, parts, and categories. *Journal of Experimental Psychology* 113(2). 169–193.
Ullmann, Stephen. 1963. Semantic universals. In Joseph H. Greenberg (ed.), *Universals of language*, 373–396. Cambridge: MIT Press.
Urban, Matthias. 2010. 'Sun' = 'eye of the day': A linguistic pattern of Southeast Asia and Oceania. *Oceanic Linguistics* 49(2). 568–579.
Urban, Matthias. 2011. Asymmetries in overt marking and directionality in semantic change. *Journal of Historical Linguistics* 1(1). 3–47.
Urban, Matthias. 2012. *Analyzability and semantic associations in referring expressions*. Leiden: Leiden University dissertation. https://scholarlypublications.universiteitleiden.nl/handle/1887/19940.

Urban, Matthias. 2022. Red, black, and white hearts: ‘heart’, ‘liver’, and ‘lungs’ in typological and areal perspective. *Linguistic Typology* 26(2). 349–374.
Van Rossum, Guido & Fred L. Drake. 2009. *Python 3 reference manual*. Scotts Valley: CreateSpace.
Wichmann, Søren, Cecil H. Brown & Eric W. Holman (eds.). 2016. *The ASJP database (version 17)*. Jena: Max Planck Institute for the Science of Human History. Available at: https://asjp.clld.org/.
Wichmann, Søren, Cecil H. Brown & Eric W. Holman (eds.). 2022. *The ASJP database (version 20)*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: https://asjp.clld.org/.
Wilkins, David P. 1996. Natural tendencies of semantic change and the search for cognates. In Mark Durie & Malcolm Ross (eds.), *The comparative method reviewed*. Oxford: Oxford University Press.
Winter, Bodo & Mahesh Srinivasan. 2022. Why is semantic change asymmetric? *Metaphor and Symbol* 37(1). 39–54.
Yu, Ning. 2002. Body and emotion. *Pragmatics and Cognition* 10(1-2). 341–367.
Zalizniak, Anna A. 2008. A catalogue of semantic shifts. In Martine Vanhove (ed.), *From polysemy to semantic change*, 217–232. Amsterdam: John Benjamins.
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 & Mikhail Voronov (eds.). 2016. *Database of semantic shifts*. Moscow: Institute of Linguistics, Russian Academy of Sciences. Available at: http://datsemshift.ru (accessed 18 February 2023).

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/lingty-2023-0032).